DOCUMENT RESUME

ED 476 146 TM 034 895

AUTHOR	Newman, Mark
TITLE	A Pilot Systematic Review and Meta-Analysis on the Effectiveness of Problem Based Learning.
SPONS AGENCY	Middlesex Univ., London (England).; Economic and Social Research Council, Lancaster (England).
PUB DATE	2003-04-00
NOTE	77p.; "On behalf of the Campbell Collaboration Systematic Review Group on the Effectiveness of Problem Based Learning."
PUB TYPE	Information Analyses (070)
EDRS PRICE	EDRS Price MF01/PC04 Plus Postage.
DESCRIPTORS	*Instructional Effectiveness; Literature Reviews; Meta Analysis; Pilot Projects; *Problem Based Learning; Teaching Methods

ABSTRACT

This paper reports on the development and piloting of a systematic review and meta analysis of research on the effectiveness of problem based learning (PBL). The systematic review protocol was pilot tested with a sample of studies cited as providing "evidence" about the effectiveness of PBL. From the 5 studies mentioned in the sample of reviews, 91 citations were identified, and 15 of these were adjudged to meet the review inclusion criteria. Of the 15, only 12 reported extractable data. All of these studies used PBL in higher education programs for health professional education. The review has established that the limited high quality evidence available from existing reviews does not provide robust evidence about the effectiveness of different kinds of PBL in different contexts with different student groups. It is apparent that there is scope for a systematic review of PBL that is specific in terms of the intervention being evaluated, comprehensive in terms of strategy used, and methodologically rigorous. One appendix contains the coding sheet used, and the other is a bibliography of reviewed papers. The second part of the paper is the review protocol. (Contains 1 figure, 10 tables, and 91 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 476 146

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

M. Newman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

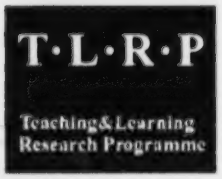
A pilot systematic review and meta-analysis on the effectiveness of Problem Based Learning

Mark Newman
Middlesex University

On behalf of the Campbell Collaboration Systematic Review Group on the effectiveness of Problem Based Learning

BEST COPY AVAILABLE

TM034895



Contents **Page**

Acknowledgements 3

Review Group membership 4

Executive summary 5

Part I: Report 9

Introduction 10

Methods 17

Objectives of Pilot Review 17

Results 19

Discussion and conclusions 29

References 39

Appendix 1: Coding sheet 44

Appendix 2: Bibliography of reviewed papers 46

Part II: Review Protocol 54

Boxes figures and tables

Box 1 Review inclusion criteria 16

Figure 1: Effect sizes with 95% Confidence intervals for category 'Accumulation of knowledge' 26

Table1: Studies excluded in preliminary screening 19

Table2: Inclusion and exclusion decisions by reviewers 20

Table3: Papers fully reviewed: decisions and reasons for exclusion 21

Table4: Curriculum design and context for included studies 34

Table 5: Reported results for experimental studies in the category 'Accumulation of knowledge' 35

Table 6: Reported results for studies using quasi-experimental designs In category 'accumulation of knowledge' 36

Table 7: Study design and reported effects in category 'improvements in practice' 37

Table 8: Study design and reported effects in category 'approaches to learning' 38

Table 9: Study design and reported effects in category 'satisfaction with learning environment' 39

Table 10: Meta-analysis: weighted mean effect sizes for category 'Accumulation of knowledge 40

Acknowledgements

The Campbell Collaboration Systematic Review Group would like to acknowledge the funding support provided for this project by the Economic & Social Research Council Teaching & Learning Research programme, The institutions of the groups members and The Learning & Teaching Subject Network Centre for Medicine, Dentistry and Veterinary Science. The group would also like to thank The Campbell Collaboration, the Cochrane Effective Practice and Organisation of Care Group, The EPPI Centre and the Department of Health Sciences at University of York for their advice and support. We would also like to thank the staff of Archway Campus Library at Middlesex University and The University of Upstate New York Medical University for their assistance in obtaining papers. We would also like to thank the people for their advice and comments on previous drafts of this manuscript including Antoinette Peters and Peter Tymms.

The report reflects the views of the members of Campbell Collaboration Review Group on the Effectiveness of Problem Based Learning and not necessarily the organizations or institutions named above or in which any of the group members are employed. The responsibility of for the content of this report lies with the review group only.

Campbell Collaboration Systematic Review group on the Effectiveness of Problem Based Learning: Membership and contribution

Name/ Dept/ Institution	Contribution to pilot review
Piet Van den Bossche Faculty of economics and business administration educational development and research, University of Maastricht, The Netherlands	Assess the quality of individual papers, feedback and commentary of analysis and report
Charles Engel Centre for Higher Education Studies, University of London, London UK	Assess the quality of individual papers, feedback and commentary on review protocol, analysis and report
David Gijbels Educational innovation and information technology (EDIT), Faculty of Law, University of Maastricht, The Netherlands	Assess the quality of individual papers, feedback and commentary of analysis and report
Jean McKendree Learning and Teaching Support Network for Medicine, Dentistry and Veterinary Medicine, University of Newcastle (UK)	Screening of identified citations for inclusion Assess the quality of individual papers, feedback on review protocol and commentary of analysis and report
Mark Newman Schools of Health & Social Sciences & Life Long Learning and Education, Middlesex University, London UK.	Develop proposal for review and obtain funding, negotiate and submit application for registration of the review, design review protocol, identify reviewers, coordinate review, identify citations from sample reviews, screen citations for inclusion, obtain copies of required papers, quality assess reviews of individual papers, data extraction, assess the quality of individual papers. Analyse included studies, write study report.
Tony Roberts South Tees Hospital Trust, North Tees Primary Care Trust and University of Durham, UK	Assess the quality of individual papers, feedback and commentary of protocol, analysis and report
Isobel Rolfe Faculty of Health, University of Newcastle (Aus)	Assess the quality of individual papers, feedback and commentary of protocol, analysis and report
John Smucny Department of Family Medicine, State University of New York Upstate Medical University, USA	Assess the quality of individual papers, feedback and commentary on protocol analysis and report
Giovanni De Virgilio Segreteria Attività Culturali, Istituto Superiore di Sanità, Rome, ITALY	Assess the quality of individual papers, feedback and commentary of protocol analysis and report

Review coordinator contact

Mark Newman
School of Lifelong Learning & Education and School of Health & Social Sciences
Middlesex University
Furnival Building
Archway Campus
2-10 Highgate Hill
London N19 5LW
Tel: 0044 (0)20 8411 6702
E-Mail: m.newman@mdx.ac.uk

Introduction

Problem Based Learning (PBL) represents a major development and change in educational practice that continues to have a large impact across subjects and disciplines worldwide. PBL is promoted by professional and funding bodies as an appropriate strategy for professional education and increasingly as the method of choice. PBL is now also now spreading into non- – professional subject areas of Higher Education. The claims made for PBL would, if substantiated, represent an important improvement in outcomes from Higher Education. Thus it is of considerable importance that questions about what forms of PBL produce which outcomes for which students in what circumstances are rigorously investigated.

There is a large volume of published work on PBL. However it is the contention of this review that despite this volume of literature, existing overviews of the field do not provide high quality evidence with which to provide robust answers to questions about the effectiveness of PBL. Systematic Reviews help by providing a comprehensive summary and synthesis of existing high quality research that may provide answers to questions and identify the areas where further primary research is needed

This paper reports on the development and piloting of a Systematic Review and meta-analysis on the effectiveness of PBL by an international group of teachers, and researchers convened under the auspices of the Campbell Collaboration.

Pilot review objectives

- To establish the evidence provided by existing published reviews about the effectiveness of PBL – defined as in increasing performance at:

❖ adapting to and participating in change;
❖ dealing with problems and making reasoned decisions in unfamiliar situations;
❖ reasoning critically and creatively;
❖ adopting a more universal or holistic approach;
❖ practicing empathy, appreciating the other person's point of view;
❖ collaborating productively in groups or teams;
❖ Identifying own strengths and weaknesses and undertaking appropriate remediation (self-directed learning)

.....when compared to other non-PBL teaching and learning strategies?

- To establish the need for a full systematic review of the effectiveness of Problem Based Learning

- To establish the value of the method of systematic review used

- To identify and clarify any problems with the review protocol, process and instruments

Method

This design of the review protocol used as a model the approach used by the Cochrane Effective Practice and Organisation of Care Group and guidelines on Systematic Reviews emerging from the Campbell

Collaboration methods group. The key principles of such reviews are that the process for identification, selection, inclusion, and synthesis of individual studies is systematic and transparent. The planned process of the review was formulated in a review protocol that specifies the review questions, the searching process, the criteria for the selection of studies for inclusion in the review, the quality criteria for assessing the quality of the individual studies and the process of synthesis. The review limits the studies that will be included to high quality experimental or quasi – experimental designs. The focus of the review is on post – school education. The inclusion criteria for 'type of intervention' were a cumulative integrated curriculum, a learning simulation format that allows free enquiry, small groups with either faculty or peer tutoring and an explicit framework followed in tutorials.

The Systematic Review Protocol was piloted using a sample of studies cited as providing 'evidence' about the effectiveness of PBL in five previous 'reviews'. These studies were all reviewed and decisions made about their inclusion in the pilot review based on the methodological criteria used in the review protocol. In practice the 'type of intervention' criteria could not be applied as in the majority of papers reviewed provided insufficient description to allow any judgement to be made against these criteria. Data were extracted from studies meeting the inclusion criteria. Where studies reported multiple effects only those that met the review criteria were included. A narrative synthesis was carried out of the included studies. A pilot Meta-analysis was carried out on a sub set of the included effects that were categorised under the heading 'accumulation of knowledge'. This was carried out using Meta-Stat software to estimate a mean effect size. Sensitivity analysis was carried out illustrate possible moderating effects of variables such as study design and assessment format.

Results

91 citations were identified from the five reviews. Of these 15 were adjudged to meet the Review inclusion criteria. Of the 15 only 12 reported extractable data. The included studies reported a range of effects that were grouped under headings discussed below. Not all of the effects reported in the included studies were included in the pilot review, only those which met the quality criteria.

The studies all reported on PBL used in Higher Education programmes for health professional education at both pre- and post- registration levels. The majority of students were in medicine and the majority of these studies reported on pre-registration medical education. Very little information was given in the papers from which data was extracted about the design, preparation or delivery processes of either the PBL intervention or the control to which PBL was being compared. Four of the included studies used a randomised experimental design, two a quasi-randomised experimental design and the remainder were controlled before and after studies. Only one study reported standard deviated effect sizes.

The effects grouped under the heading 'improvement on practice' all used different outcomes and measurement instruments. In only one of the studies was sufficient data provided to calculate effect sizes. This makes it difficult to synthesise the study results. One study reported attitudes to practice and found effect sizes that favoured PBL. Another measured nursing process skills and of the seven effects reported five favoured the control group. The third study reported consultation skills and on all the effects reported the results favoured the control group. However in this study which used a quasi – experimental design the nature of the control group intervention and the outcome measures used would appear to have put the control group at a distinct advantage.

Two studies reported effects on 'approaches to learning'. The two studies used different instruments and reported a total of five effects. In both studies the results favoured PBL on all the scales. The PBL groups had less of the undesirable and more of the desirable approaches to learning after the intervention. However it is interesting to note that the overall picture was deterioration in the approaches to learning of both the PBL and control groups which the PBL appears to have been mitigating.

In only one of the included studies did the effects reported on 'satisfaction with the learning environment' meet the review inclusion criteria. The study, set in an undergraduate medical education programme, required students to rate their experience on a series of scales (effects). On all except two of the nine effects reported the effect size favoured the PBL group.

The majority of effects reported could be grouped under the heading 'accumulation of knowledge'. Reported effect sizes ranged from d= -4.9 to d=2.0. There were sufficient effects reported to pilot a meta-analysis. The meta-analysis included 14 effects reported in eight different studies. The mean effect size was d= -0.3 but the 95% confidence interval did not exclude a positive effect. Sensitivity analysis (see table S1 below) suggested that study design, randomisation, level of education and assessment format are all potential moderating variables. Importantly the 95% confidence intervals in many of the sub group analysis do not exclude potentially 'large' effect sizes of d = + 1.0 or – 1.0. An effect size of d = 1.0 would mean that 84% of students in the control group were below the level of the average person in the PBL groups. An effect of this magnitude would appear to have important practical significance.

Table S1: Meta- Analysis: Weighted mean effect sizes for outcome 'knowledge' total and sub groups

Moderator	Grouping	Mean Effect size	St. Dev.	N	95% C.I
	Outcome-Knowledge	-0.3	15.81	1904	-1.0 to 0.4
Study Design	Experiment	-0.4	16.47	1719	-1.1 to 0.4
	Quasi – experiment	0.6	6.99	185	-0.4 to 1.6
Randomisation	Random	-0.8	24.63	757	-2.6 to 1.0
	Non – random	0.1	3.77	1174	-0.1 to 0.3
Assessment format	MCQ	-0.3	16.79	1676	-1.1 to 0.5
	Written assessment	0.3	3.4	228	- 0.1 to 0.74
Qualification of student	Pre – qualification	-0.4	16.56	1700	-1.1 to 0.7
	Post – qualification	0.5	6.7	204	-0.4 to 1.4

Discussion and Conclusions

The pilot Systematic Review has established that the limited high quality evidence available from existing reviews does not provide robust evidence about the effectiveness of different kinds of PBL in different contexts with different student groups. It is apparent that there is scope for a systematic review of PBL that is specific in terms of the 'intervention' that is being evaluated, comprehensive in terms of strategy employed to identify potential evidence and methodologically rigorous in terms of the criteria used to evaluate the quality of evidence.

The pilot review demonstrates the potential value of a Systematic Review and meta-analysis in summarising and synthesising existing research to begin to provide robust answers to questions of effectiveness and to identify issues for further primary research. The pilot review also demonstrated that the Systematic Review approach taken by the Cochrane Effective Practice and Organisation of Care group can be successfully applied in a purely educational context.

However, the pilot review also highlighted a number of conceptual, methodological and practical problems that will need to be addressed by a full review and by those interested in PBL. The reporting of studies of education interventions that are labelled 'PBL' by the authors does not in general appear to contain sufficient description of either the experimental or control interventions. This makes it difficult to distinguish between different types of PBL and even to distinguish between PBL and other educational interventions. In part this is an issue that can be addressed by Journal Editors and study authors adhering to agreed guidelines in the reporting of studies. However, whereas such guidelines exist for the reporting of methodological aspects of studies designs no such guidelines exist for describing educational interventions. This is not just a technical issue but also conceptual. For example PBL is often described as a 'philosophy', the question arises of how one might meaningfully describe the particular philosophy of learning used in a particular programme. Such questions are not only relevant for systematic reviews but also to primary research and theory about PBL. Whilst there have been some useful attempts to provide descriptive criteria for PBL programmes such as those proposed by Howard Barrows (2000), these do not appear to be widely used. Even where such criteria could be agreed and used by the various PBL communities their retrospective application is likely to prove difficult and time consuming, requiring the use of additional sources of information other than a single journal article. The pilot review therefore indicates that that the resources required to conduct a full review within a reasonable timescale will be significant.

However none of these difficulties is insurmountable and the experience of the pilot review demonstrate that collaboration and cooperation across countries and disciplines is possible is a fruitful experience for those involved and generates valuable knowledge for teachers and researchers alike.

Part I: Report

Problem Based Learning

Problem Based Learning (PBL) represents a major development and change in educational practice that continues to have a large impact across subjects and disciplines worldwide. There has been a steady growth in the number of programmes and institutions that have adopted PBL around the world. This transformation has been encouraged by an almost evangelical PBL movement that has published a wealth of anecdotal material extolling the virtues of PBL (Wilkie 2000). PBL has been endorsed by a wide variety of national and international organisations. These include the Association of American Medical Colleges (Muller 1984) the World Federation of Medical Education (Walton & Matthews 1989), The World Health Organisation (1993), World Bank (World 1993)and the English National Board for Nursing Midwifery and Health Visiting (English National Board 1994). However it is not always clear what exactly is being done in the name of PBL (Maudsley 1999). There are also a growing number of references in the literature to 'adapted' or 'Hybrid' PBL courses and courses called 'Enquiry' or 'Inquiry' Based learning which are apparently based on but not the same as Problem Based Learning (Savin-Baden 2000).

What is Problem Based Learning ?

There is no single unanimous position about the theoretical basis for or practice of PBL. The philosophical and theoretical underpinnings of PBL were not explicit in the early PBL literature (Rideout & Carpio 2001). Barrows, a pioneer of PBL, explains that the he and the other developers of the original the McMaster PBL curriculum had no background in educational psychology or cognitive science. They just thought that learning in small groups through the use of clinical problems would make medical education more interesting and relevant for their students (Barrows 2000). Historically the development of PBL in medical education appears to have been heavily influenced by Cognitive Psychology (Norman & Schmidt 1992;Schmidt 1983;Schmidt 1993). More recently and has PBL has expanded into other disciplines theoretical justification has also been derived from other 'educational' theorists who place emphasis on different aspects teaching and learning such as Dewey (1938) and participation; Schon (1987) and Reflective Practice; and Vygotsky (1978) and the communal social construction of learning.

A review of the field, found that the practice of PBL was described in a variety of ways that could be summarised as a complex mixture of general teaching philosophy, learning objectives and goals and faculty attitudes and values (Vernon D.T & Blake 1993). Walton and Matthews (1989) argue that PBL is to be understood as a general educational strategy rather than merely a teaching approach. They present three broad areas of differentiation between PBL and the 'traditional' subject centered approaches

1. Curricula Organisation: Around problems rather than disciplines, integrated, emphasis on cognitive skills as well as knowledge.
2. Learning environment: use of small groups, tutorial instruction, active learning, student centered, independent study, use of relevant 'problems'.
3. Outcomes: Focus on skills development and motivation, abilities for life long learning

Engel (1991)focuses on curriculum design as a major area of difference. He describes the essential characteristics of problem-based curricula as cumulative (repeatedly reintroducing material at increasing depth) integrated (de-emphasising separate subjects), progressive (developing as students adapt) and consistent (supporting curricula aims through all its facets). Barrows (1986) differentiates between six types of PBL by method. Savin-Baden (2000) identified five models of PBL in operation in different curricula. She argues that the important differentiation is the way that knowledge, learning and the role of the student are conceptualised and manifest in the curricula. Many accounts of PBL emphasise the

steps described by Schmidt (1983)are:

1) clarifying and agreeing on working definitions of unclear terms/concepts;
2) defining the problem(s), agreeing which phenomena require explanation;
3) analysing components, implications, suggested explanations (through brainstorming) and developing working hypothesis
4) discussing, evaluating and arranging the possible explanations and working hypotheses
5) generating and prioritising learning objectives
6) going away and researching these objectives between tutorials
7) reporting back to the next tutorial, synthesising a comprehensive explanation of the phenomena and reapplying synthesised newly acquired information to the problem(s)

Assessing the impact of educational interventions the role of Systematic Reviews

Whilst the principle of research reviews is well established in education the appropriate process and purpose of such exercises is contested (Schwandt 1998). This contest is in large part linked to an ongoing debate about the methods used to generate knowledge about appropriate and effective educational practices which in turn is linked to debated about the role, purpose and nature of education (Oakley 2003). The traditional view of the research review is that it seeks summarise what is know about a particular topic in order to inform policy, practice, debate and further research. The traditional narrative literature review has attempted to undertake this role but has been criticized for not being sufficiently rigorous in specifying or utilising an explicit methodology (Gough & Elbourne 2002).

Systematic reviews can be a valid and reliable means of avoiding the bias that comes from the fact that single studies are specific to a time, sample and context and may be of questionable methodological quality. They attempt to discover the consistencies and account for the variability in similar appearing studies (Davies & Boruch 2001). A systematic review is a piece of research in which specific methods are used to reduce distortions or inaccuracies (EPPI Centre 2000). The emerging science of systematic reviewing includes methods for locating, appraising and synthesising evidence that can be viewed as explicit attempts to limit bias (Petticrew 2001). Importantly the systematic review provides information by summarising the results of otherwise unmanageable quantities of research (Light & Pillemer 1984).

There is a consensus emerging about the need for systematic reviews covering selected topics in medical education (BEME 2000). Such reviews will identify the existing evidence, provide at least some answers to the review questions and/or will provide directions for future primary research (Wolf 2000). A relevant example is the reviews of the effectiveness of continuing medical education carried out by the Cochrane Effective Practice and Organisation of Care Group that have been useful in identifying formal educational practices which appear ineffective (Davis & Thomson M.A 1995).

The development of the science of systematic reviewing has been underway since the 1960's but has come to prominence more recently through the work of the International Cochrane Collaboration that has developed a framework for the conduct and dissemination of systematic reviews in healthcare (http://www.cochrane.org/). The systematic reviews produced by the Cochrane Collaboration are driven largely by clinicians that are seeking high quality of evidence for clinical decision making. The bias minimisation advantages of the properly conducted Randomised experiment in conjunction with its desirable inferential properties lend themselves ideally to this type of question (Egger et al. 2003), hence most Cochrane Collaboration groups have limited their reviews to this kind of primary study (Petticrew 2001). However, there maybe contexts in which randomised experiments are not feasible in which case researchers will use other designs. The use of quasi- experimental designs seems to be common in

educational research and the Cochrane Effective Practice and Organisation of care group (EPOC) also includes high quality studies using these designs within its reviews.

The rationale for a systematic review of Problem Based Learning

Any researcher wishing to investigate the effectiveness of any educational intervention is confronted by the problems created by the disorganisation (i.e. spread over many different journals, books and databases) and volume of the literature. PBL has arguably been one of the most scrutinised (i.e. researched) innovations in professional education (Maudsley 1999). A simple illustration of this is that a search of the Medline bibliographic database on line via PUBMED using the search terms 'Problem Based Learning' in Winter 2002/3 yields a reference list of over 1000 citations. Even in this comparatively well indexed database a large proportion of these references will not in fact be about Problem Based Learning and a proportion of references to papers on Problem Based Learning that are on the bibliographic database will not be retrieved using these search terms. The Medline bibliographic database covers only journals about or relevant to healthcare. Thus education journals and the journals of other subjects and disciplines are not covered. And yet these all are possible publication outlets for studies of Problem Based Learning. A brief search using the terms Problem Based Learning produced 804 'hits' on the Science Citation Index, and 384 in the Social Science Citation Index[1]. A Systematic Review would be required to identify and synthesise this evidence, unless there are existing good quality and up to date reviews that can provide empirical answers to the question (Glanville & Sowden 2001).

There have been at least five 'reviews' of PBL that have attempted to provide evidence about the conditions and contexts in which PBL is more effective than other educational strategies. A major potential weakness of these reviews is that they include, with one or two exceptions, only studies of PBL in the education of health professionals. Three of the reviews were published in the same journal in the same year (Albanese & Mitchell 1993;Berkson 1993;Vernon & Blake 1993). These three reviews, which are perhaps the most well known, are difficult to interpret due to the lack of clarity about the review methods used and apparent differences in approach between the reviews. The reviews include primary studies with different designs and of differing quality (Wolf 1993). Of the citations identified by the review authors as providing 'evidence' about PBL only eight appear in all three reviews, whereas 49 citations appear in only one out of the three.

The criteria for inclusion of studies in a 'Meta-analysis' of PBL carried out by Van Den Bossche and colleagues (2000) are explicit. However the study design and quality criteria applied to the primary studies appear to be fairly minimal, raising the possibility that studies with significant weaknesses in terms of bias minimisation have been included in the review. The authors recognised the risk of bias in the location of studies and described, by the standards of most reviews, a fairly comprehensive search strategy. However the search included only a limited number of Bibliographic Databases (not including MEDLINE) and the search strategy only a limited number of terms and would therefore also appear to be inadequate in these respects (Egger & Smith 1998).

Smits and colleagues (2002a) carried out a review of the effectiveness of PBL in continuing medical education. An explicit search strategy including a wide range of bibliographic databases was used but it appears that limited attempts were made to locate the so-called 'grey' literature. This review adopted strict methodological inclusion criteria by including only randomised and controlled trials. Whilst this will have

[1] February 2003 via WWW using Ovid interface

reduced the risk of bias in the individual studies (Cook & Campbell 1979) it may also have meant that potentially useful studies of PBL using other designs were excluded.

The reviews all provide only limited descriptive information about the educational interventions that are called Problem Based Learning or the interventions to which PBL is compared. Unsurprisingly the reviews referred to above came to differing conclusions. Vernon and Blake (1993) concluded "results generally support the superiority of the PBL approach over more traditional academic methods". Albanese and Mitchell (1993) whilst acknowledging the weaknesses of the research literature concluded that PBL was more nurturing and enjoyable and that PBL graduates performed as well and sometimes better on clinical examinations and faculty evaluations. However, they also concluded that PBL graduates showed potentially important gaps in their cognitive knowledge base, did not demonstrate expert reasoning patterns, and that PBL was very costly. Berkson (1993) was unequivocal in her conclusion that "the graduate of PBL is not distinguishable from his or her traditional counterpart". She further argued that the experience of PBL can be stressful for the student and faculty and implementation may be unrealistically costly. The two more recent reviews also came to differing conclusions. Van Den Bossche and colleagues (2000) concluded that PBL had a positive robust effect on the skills of students but a negative non-robust effect on knowledge. The review by Smits and colleagues (2002a) concluded that there was no consistent evidence that PBL is superior to other educational strategies in improving doctors knowledge and performance. The reviews themselves therefore provide contradictory evidence about the effects of different kinds of PBL in different learning contexts.

The Establishment of the PBL review group

A Systematic Review was proposed as part of the Project on the Effectiveness of Problem Based Learning (PEPBL - http://www.hebes.mdx.ac.uk/teaching/Research/PEPBL/index.htm.). This project seeks to investigate Problem Based Learning from the perspective of the potential user of PBL seeking to decide whether or not to use a PBL curriculum. From this perspective the purpose of a review is to investigate the evidence for the relative costs and benefits of using PBL, as opposed to any other teaching and learning approach. The PEPBL project developed an international network of practitioners, policymakers and researchers interested in PBL. This network was used to recruit volunteers to collaborate in the review process. Due to resource constraints members of the review group were required to have sufficient 'ability' to review a study with potential for inclusion using the materials provided with no additional support. The members of the review group who participated in the pilot Systematic Review are listed at the front of this report.

The PEPBL project originates in health care education and therefore the work of The Cochrane Collaboration and in particular the Cochrane Effective Practice and Organisation of Care Group (EPOC 1998) is familiar to those involved in the project. The EPOC group was approached as a natural home for the review but felt that the fact that a review would go beyond the boundaries of post-graduate health care education meant that they could not provide logistical support for it. At about this time the Evidence Informed Policy and Practice in Education Initiative funded by the Department for Education & Employment (DfEE) was being established at the EPPI Centre in the UK. Discussions were held with the EPPI Centre about the possibility of establishing a PBL review group within this initiative. However, the conditions of the DfEE funding (limited to school aged education at that stage) and the fact that the PBL review group wanted to take a different approach to that used by the EPPI Centre meant that the review could not be accommodated within the EPPI network. The Best Evidence in Medical Education (BEME 2000) is another emerging collaboration linked with the Association for the Study of Medical Education (ASME). The focus of this group is identifying evidence in relation to medical education (BEME 2000). This group has not selected PBL as one of its review priority areas and is still establishing it practical and methodological approach to reviews.

The goal of the Campbell Collaboration is to produce disseminate and continuously update systematic reviews of studies of the effectiveness of social and behavioural interventions including education interventions. It is described as the sister Organisation to the Cochrane Collaboration and 'leaning heavily on its shoulders' for its model of infrastructure, development and processes (Boruch et al. 2001) The Campbell Collaboration was inaugurated in 2000 and its infrastructure and methodology is still under development. The PBL review group developed a protocol that was submitted and accepted by the Campbell Collaboration Education coordinating group by the end of 2001.

Systematic Reviews can be completed without the support of information professionals, but their researchers are likely to produce searches that are less sensitive, less specific and to do so more slowly, particularly when searches have to be carried out across subject/ disciplinary boundaries (Dickersin et al. 1994). The review proposal envisaged that this support would be obtained as a result of registering with a Cochrane collaboration group. As reported above this did not transpire and as yet the Campbell Collaboration do not have the resources to provide such support to review groups. Such support was eventually identified through one of the review group collaborators. However due to a change in institutional priorities the review collaborator and information professional had to withdraw from the

review. At that time (early 2002) the review group decided that it would be more productive to go ahead with a pilot review rather than to continue to wait for professional information support.

Review questions

As has been already been noted there is no universal agreement about the goals or aims of PBL. Engel (1991) argues that where PBL is adopted one of the aims is to assist students towards achieving a specific set of competencies, that will be important to them throughout their professional life, irrespective of the profession in which they will come to practice. The competencies are suitably broad to encompass a range of interpretations and thus were used to provide a heuristic framework for the review questions. The initial review questions are as follows. Does PBL result in increased participant performance at:

- adapting to and participating in change;
- dealing with problems and making reasoned decisions in unfamiliar situations;
- reasoning critically and creatively;
- adopting a more universal or holistic approach;
- practicing empathy, appreciating the other person's point of view;
- collaborating productively in groups or teams;
- Identifying own strengths and weaknesses and undertaking appropriate remediation (self-directed learning)

.....when compared to other non-PBL teaching and learning strategies?

The approaches taken to the operationalization and measurement of student performance in each these areas are likely to vary between PBL curricula. All reported effects that meet the inclusion criteria will be included in the review. The seven goals identified above will be used as a framework for analysis and synthesis of the findings from individual studies. If possible (i.e. the data allow) a secondary review question about whether an 'authentic' PBL curriculum delivers a greater improvement in performance (as defined above) than so called 'hybrid' curricula will be carried out.

Review Design

The review protocol (see part II) gives details of the methods used to identify, assess the quality of and synthesise the included studies. The review question(s) are specifically concerned with the effectiveness of PBL in comparison with other teaching and learning strategies and as noted earlier the overall perspective taken is that of the research user confronted by a decision about whether or not to use PBL. Whilst it maybe the case that such a review could include primary research studies that had used a wide variety of designs, different designs engender different patterns of threats to internal validity and thereby permit causal inferences with different levels of certainty. The true experimental design is considered most useful to demonstrate programme impact if randomisation in the assignment of treatment can be met (Boruch & Wortman 1979). The experiment is a particularly efficacious design for causal inference. Random assignment creates treatment groups that are initially comparable (in a probabilistic sense) on all subject attributes. It can then be concluded that any final outcome differences are due to treatment effects alone, assuming that other possible threats to validity have been controlled (Tate 1982).

For this reason the review design followed the approach used by The Cochrane Collaboration in which Randomised experimental designs are considered the 'gold standard'. However there are numerous reasons why a review should consider other designs. Firstly Randomised experiments are more plentiful in

some fields. Secondly non–randomized studies may provide information that has not been provided in Randomised studies. Thirdly both Randomised experimental designs and non-experimental designs vary enormously in quality. The results of poor quality Randomised experiments may be less helpful than better conducted quasi-experimental studies (Shadish & Myers 2002). The design of the review protocol, data extraction tools and overall review process used was therefore derived from the guidance for reviewers produced by The Cochrane Effective Practice and Organisation of Care Review Group (EPOC 1998). This Cochrane group includes quasi- experimental research designs in their reviews.

Box 1: Summary of PBL Review minimum inclusion criteria

- The review will only include participants in post-school education programmes.

- Study designs included:

Randomised Controlled Trials, Controlled Clinical Trials (CCT, Interrupted Time Series (ITS), Controlled Before & After studies (CBA). Qualitative data collected within such studies e.g. researchers observations of events will be incorporated in reporting. Studies that utilise solely qualitative approaches will not be included in the review. For each study design a set of minimum quality criteria is used.

- Methodological inclusion criteria

The minimum methodological inclusion criteria across all study designs are the objective measurement of student performance/behaviour or other outcome(s). (Blinding, reliability, follow-up)

- Type of Intervention

The minimum inclusion criteria for interventions for the initial review are:
Cumulative integrated curriculum, Learning via simulation formats that allow free enquiry (i.e. not problem solving learning), Small groups with either faculty or peer tutoring, An explicit framework is followed in tutorials e.g. Maastricht 7 steps.

The criteria given in box 1 will be used to select studies for inclusion in the review. More specifically only effects (i.e. particular outcome measures) that meet the quality criteria will be included. Deciding on inclusion criteria for the type of intervention is in effect equivalent to deciding what is the cut off point at which an educational intervention can no longer be considered to be PBL. This would seem to be incongruent with the Systematic Review's aim to explore the costs and benefits provided by different types of PBL. However, for practical reasons alone the boundaries have to be located somewhere, otherwise every learning intervention would be eligible for inclusion in the study. The review group decided on the basis of their knowledge of the PBL literature, that to be eligible for inclusion interventions would as a minimum need to meet the four inclusion criteria outlined in box 1. This should not be interpreted as an argument that any thing that meets these criteria 'is PBL' whilst anything that does not is not, but rather a pragmatic response to a practical problem posed by reviewing methodology. Studies that meet the research methodology criteria and more than one but not all of the above criteria (i.e. maybe considered a hybrid or a combination PBL curriculum) will be included in the database for analysis of the secondary review question.

The PBL review group succeeded in producing a review protocol that was registered with the Campbell Collaboration. This is in itself an important development not only for research into PBL and for the development of this style of Systematic Review in education more generally. The total amount of resources required for a full Systematic Review is essentially unknown but proved ultimately to be beyond the scope of resources that could be mobilised within the three years funding of the PEPBL project. Whilst some sections of the PBL and education community appear to accept the need for and understand the value of Systematic Reviews of the type proposed it was felt that the argument would be more persuasive if an example of the potential of such a review could be provided. The review group therefore decided to conduct a pilot review with the following objectives:

- To investigate what 'high quality' evidence about the effectiveness of PBL compared to any other teaching and learning strategy can be derived from a selected group of existing 'reviews'
- To establish the need for a full systematic review of the effectiveness of Problem Based Learning
- To establish the value of the method of systematic review used
- To identify and clarify any problems with the review protocol, process and instruments

Pilot Sample

For the purpose of the pilot study the sample are those papers cited in the in the five 'review' papers referred to earlier as providing evidence of the effectiveness of PBL (Albanese & Mitchell 1993;Berkson 1993;Smits et al. 2002b;Van den Bossche et al. 2000;Vernon D.T & Blake 1993).

Review Process

The pilot review followed the design and methods outlined in the review protocol with the exception that no searching was undertaken. The review co-ordinator examined the reviews and identified the relevant citations. For each citation either an abstract or full text copy was obtained. Two members of the review team screened the citations by reading through all the abstracts and or full papers to eliminate those that obviously did not meet the minimum inclusion criteria (see table 1 for a list of citations excluded on screening). As planned only studies that met all of the criteria shown in box 1 were to have been included in review. However it became apparent during course of the review that very few papers provided sufficient description to allow decisions to be made about whether they met the 'type of intervention' inclusion criteria. In practice therefore only one citation was excluded on the basis of not meeting these criteria. In this case the authors had called their intervention something completely different (Vu & Galofre 1983)

Full copies were obtained of each of the remaining papers and these were then distributed amongst the reviewers for quality appraisal and data extraction. Each paper was reviewed independently by two reviewers. The allocation of paper to reviewers followed three principles. Firstly that the reviewers should have no connection with the institutions in the study being reported. Secondly the same reviewers should review the papers reporting studies carried out at the same institution. Thirdly each person should be the second reviewer for every other member of the review panel at least once. The exceptions to this were the

papers written in Dutch that were reviewed by the two Dutch language-speaking members of the review group.

Where there were differences of opinion between the first two reviewers the article was passed to a third member of the panel for independent review. At all stages the process used and outcomes are explicit. A full list of included and excluded studies is provided to allow for independent scrutiny of the review process. The completed quality assessment and data extraction tools were returned to the review co-ordinator who lead the process of producing a report of the review analysing and synthesising the results where appropriate

RESULTS

Ninety one citations were identified from the five reviews (see appendix two for a full bibliography). The screening process eliminated 60 citations that obviously did not meet the review inclusion criteria (See table 1). Of these 60, 43 were excluded as the study design did not meet the inclusion criteria, eleven were discursive overview papers, 4 were not reporting studies of the effectiveness of PBL and in one the subjects were high school students.

Table 1: Studies excluded in preliminary screening

Paper author	Reason for exclusion
Woodward & Ferrier, 1982:	Single group post test design
Woodward & Ferrier, 1983:	Single group post test design
Woodward, McAuley, & Ridge 1981:	post –test only design
Woodward, C 1990:	Post- test only design
West, & West, 1987:	Single group post test design
West, Umland, & Lucero, 1985:	Single group post test design
Vu & Galofre, 1983:	Not PBL Objective Based mastery programme
Vernon, Campbell & Dally, J. C. 1992:	Post test only
Van Hessen & Verwijen, 1990:	Post test only design
Van Aalst et al 1990:	Personal reflections of student not empirical study
Tolnai, S. 1991:	Post –test only design
Son & Van Sickle 2000:	High school students
Shin, Haynes, & Johnston 1993	Post test only design
Schwartz, et al 1997	Discursive paper
Schuwirth 1998:	Post test only design
Schmidt, H. G., et al 1996	post test only design
Saunders, N et al 1990	post test only design
Saunders, Northup, & Mennin, 1985:	post test only design
Santos Gomez, L., et al 1990:	post test only design
Richards, B. et al . 1996:	post test only
Rangachari, P. K. 1991:	Single group post test design
Puett, D and Braunstein, J. J 1991:	Single group post test design
Post, G and Drop, M 1990:	Post test only design
Polglase, Parish, & Camp, 1996:	Single group post test design
Patel, Groen, and Norman 1991	Post test only design
Olson, J. O. 1987:	Single group post test only design
Nolte, Eller & Ringel 1988	Single group post –test only design
Newble & Gordon 1985:	About the learning styles of medical students not PBL
Newble & Clarke 1986	post test only design
Neufeld, Woodward, & MacLeod, 1989:	Discursive overview of McMaster PBL programme some evaluation studies mentioned no data
Neufeld & Sibley, 1989:	Discursive paper
Neame, R 1989:	Descriptive
Mitchell, R 1992:	About development & testing of outcome measuring instrument not impact of PBL
McAuley & Woodward 1984:	Single group post test design
Maxwell & Wilkerson 1990:	Single group post test Design
Martenson et al 1985	Comparison of achievement of PBL students with average achievement prior to PBL programme
Klass, D et al 1987:	About method of assessment not PBL
Kassebaum, Averbach, & Fryer 1991:	Within subject design no washout, no objective measures of performance
Imbos, et al 1984:	About Maastricht progress test. In one example able to compare results with other school non PBL But no data or detail given only graph and narrative
Hmelo,Gotterer & Bransford 1997:	Post test only design
Hamad, B. 1985:	Discursive paper
Gordon, M. J 1978:	not PBL
Goodman et al 1991:	post test only design
Eisenstaedt 1990	Post –test only design
Finch 1999	post test only design
Drop, & Post 1990:	Single group post test design
Distlehorst & Robbs 1998:	Post test only design
Dietrich et al 1990:	Single group post -test design
Des Marchais, et al 1992:	Discursive paper
De Vries, Shmidt H, & and De Graf, 1989	Discursive overview of other Maastricht studies
Colditz, G. A. 1980:	Single group post test design
Clarke, Feletti, & Engel1984:	Single group post test design
Claessen & Boshuizen, 1985	Post test only design
Boshuizen & Schmidt 1993:	post test only design
Blumberg & Michael 1992	post test only design
Bickley, et al 1990:	Comparison only with national average
Barrows & Tamblyn 1977:	Discursive
Anderson, Camp, & Philip 1990	Post test only design
Albano et al 1996	Post test only design
Al Haddad & Jayawickramarajah 1991	Single group post test only design

The 31 remaining citations were distributed amongst the review panel for quality assessment and data extraction. The list of papers reviewed and the reviewers decisions on inclusion are give in table three. The two reviewers came to the same conclusion on whether to include or exclude a particular paper in 24 cases. In the seven cases where the reviewers disagreed the paper was reviewed again by a third member of the team and the majority decision accepted. Table three provides a breakdown of the inclusion/ exclusion rate for each reviewer. The figures indicate the inclusion rates varied between reviewers ranging from 30% to 87%. Taken alone this may indicate that reviewers were applying the inclusion criteria differently. However, the figures for the proportion of times an individual disagreed with the second reviewer of the same paper is consistently fairly low which suggests that the differences between the reviewers is more likely to be explained by distribution of papers reviewed rather than differential application of the review criteria.

Table 2: Inclusion and exclusion decisions by reviewer

Reviewer	Include	Exclude	Different to 2nd reviewer
JS	7 (77%)	2 (23%)	3 (33%)
IR	6 (66%)	3 (33%	2 (22%)
GdV	3 (42%)	4 (58%)	2 (28%)
DG	1 (50%)	1 (50%)	0
PvB	3 (42%)	4 (58%)	0
MN	3 (30%)	7 (70%)	2 (20%)
TR	4 (44%)	5 (56%)	4 (44%)
JM	7 (87%)	1 (13%)	1 (13%)

Five of the reviewed papers reported on the 'New Mexico Experiment' that evaluated the Primary Care Curriculum (PCC) that used Problem Based Learning. (See table 3 for details of papers). These papers were reviewed as a block by the same two reviewers who concluded that the 'New Mexico Experiment' should be included in the review. Some of the papers report the same data and some of the papers report data on a range of different outcome measures that were used. Only the paper by Mennin et al (1993) provides data in sufficient detail for extraction. Similarly, four papers report results from the evaluation of the 'New Pathway Programme' at Harvard Medical School (see table three for details of the papers). Two reviewers also reviewed these papers as a group. The reviewers concluded that the 'New Pathway Programme' experiment should be included in the review. Some of the cited papers were descriptive and others reported the same data. The data from one paper was extracted (Moore et al.1994).

The reviewers were required to undertake data extraction for studies that they felt should be included. The review coordinator then synthesised the two reviewer's reports. During this process it became clear that even where the reviewers agreed about inclusion of the study in some cases they had interpreted the study and/or the review inclusion criteria differently. The technical and practical issues that account for and result from this will be discussed further below. The main area of difference between review team members was over the issue of response or follow-up rates. The review protocol follows the approach taken by Cochrane EPOC reviews in that the minimum acceptable level of response or follow-up rate is set at 80%. It became apparent that not all reviewers had applied this criterion and in subsequent discussion felt that the 80% figure was too high. For those studies finally included data for a particular effect was only reported where the response rate is less than 80% if the assessment was both blinded and reliable.

Table 3: Papers fully reviewed: Decision and reason for exclusion

Paper Author	Include (I)/ Exclude (E)	Reason for exclusion
Antepohl W,	I	
Baca E,[1]	E	New Mexico duplicate
Benjamin EM,	E	Study of the effect of guideline implementation rather than PBL
Block & Moore G. [2]	E	Description of study methods
Block et al 1993 [2]	E	Unpublished paper – data later published in Moore et al 1994
Blumberg, P	E	No reliability reported – response rates lower than <80%
Chan	I	
Bridgham R	E	Not comparing PBL with another curriculum
Coles CR.	I	
Donner RS	E	Retrospective analysis using data not collected for research purposes
Doucet MD,	I	
Farquhar L	I	
Grol R	I	
Heale J et al	I	
Hmelo CE.	I	
Imbos 1982	E	Anecdotal narrative account of PBL experience
Jones, J	E	Comparative post test only design
Kaufman A, [1]	E	Duplicate New Mexico
Lewis KE	I	
Mennin S, et al 1993[1]	I	
Mennin SP, & Friedman M [1]	E	Duplicate New Mexico
Moore-West M, [1]	E	Description of evaluation methods
Moore GT. 1991 [2]	E	Retrospective single subject A/B/A design, no washout , no randomization, no control for order effects
Moore GT et al 1994 [2]	I	
Moran JA, Kirk P,	E	Non equivalent control group, non contemporaneous data collection
Morgan HR.	E	Comparative post test only design
Premi J	I	
Tans 1986	I	
Verhoeven B.H.	I	
Verwijnen M,	I	
Woodward CA	E	Comparative post test only design

1: Papers reporting on Primary Care Curriculum at University of New Mexico
2: Papers reporting on New Pathway programme at Harvard University

There were three cases where the reviewers recommended that studies should be included but from which it was not possible to extract data. In the case of the paper by Hmelo (1998) there was a lack of clarity in the paper about the composition of the intervention and control groups and no data are reported (only graphs and statistical test results). For the paper by Verwijnen and colleagues (1990) the reviewers disagreed about the inclusion of this study which reports outcomes from the undergraduate medical programme at Maastricht medical school. The basis of this disagreement was about the design of the study an issue that is discussed in more detail below as the issue concerns all the Maastricht studies. The reviewer agreed with the inclusion verdict, however there was no data in the paper to extract. The study by Heale and colleagues (1988) reports the findings of a randomised controlled trial in which Problem Based Learning in small groups is compared with Problem Based learning in large groups and traditional didactic learning. However the reporting of the results is somewhat unclear with regard to response rates in each group, construction of the assessment instrument and results at each assessment. In a full review the authors would be contacted for clarification and copies of the original data.

Synthesis of included studies: Key characteristics (practical heterogeneity)

Descriptive information about each of the included studies is given in table four[2]. As noted earlier in the report in practice it proved to be largely impossible to make decisions about inclusion or exclusion using

[2] The size and design of tables four to nine precluded their inclusion in the main body of the report. They can be found at the end of part 1 of the report. The codes for the tables can be found in Appendix 1

the 'type of intervention' criteria. Only one study was excluded on these grounds. The included studies should not therefore be assumed to meet these criteria. There is generally more description in the reports of continuing education than the reports of undergraduate studies. This is not to say that this information could not be obtained elsewhere. There is a book describing the New Pathway Programme for example (Armstrong et al. 1994). Clearly in a full review such information could be obtained but this could have significant implications for the resourcing of any full review.

The subjects in all of the studies are trainee or qualified health professionals and with only two exceptions (nurses and physiotherapists) the studies concern the education of doctors. All the studies were carried out in North America and Western Europe. All the included studies featured University delivered programmes. With respect to motivation of participants there appear to be no specific incentives for any programmes apart from the desire to graduate in undergraduate programmes and the requirement to obtain Continuing Medical Education accreditation points. The PBL contact time and curriculum timespan ranged from a total of six hours contact time over three months to two and a half days per week for two years. The training of tutors was mentioned only in the New Mexico study (Mennin et al. 1993) and in the Headache management study (Premi et al. 1994). Only one paper mentioned the design of 'triggers' or 'problems' (Tans et al. 1986) and no papers discussed the use of standardised patients or student grading. In one study the PBL curriculum was the control group for the assessment of another method of education and is was described in much detail (Grol et al. 1989).

Given the lack of detailed information in the papers included it is difficult to distinguish where PBL is used as the only curriculum approach and where it is one of a range of philosophies and/or methods used. It appeared that all of the studies used a mixture of PBL and other approaches. In the study by Tans and colleagues (1986) the outcomes are reported for students assessments on one PBL 'only' module but this is the only PBL module in an otherwise non-PBL programme. The New Mexico curriculum is based on a philosophy that combines small group problem based learning with 'rural based education' (Kaufman et al. 1989). In the study reported by Chan and colleagues (1999), General Practitioners were randomised into either a PBL internet group that received web based case material to discuss as group by e-mail or a group that were just given the web based material without the interactive group discussion. The subject of learning was very specific (depression in the elderly) and the learning outcome appears to have been limited to improving the students knowledge of the subject matter. This could be interpreted as taking the intervention as without the remit of PBL.

Synthesis of included studies: Study quality (methodological heterogeneity)

The study design and methods used in the included studies are summarised in tables five to nine. Of the 12 studies from which data has been extracted, four were randomised experiments, two quasi–randomised experiments (CCT) and six used quasi-experimental designs (CBA). The quasi experiment reported by Farquar and colleagues (1986) was a Controlled Before and After study using a matched pairs design, whereby students who volunteered for either the control or intervention group were matched by Medical College Admission Test (MCAT) scores. Problem Based Learning was introduced on an experimental basis at the University of New Mexico medical education programme in the early 1980's allowing the comparison of students who took the conventional programme with students taking the programme that used PBL. The included paper by Mennin and colleagues (1993) reports results for six subgroups of students within the two programmes including two groups that were randomly allocated to either the PBL or the traditional programme. The randomised trial therefore compares students who volunteered for the PBL programme but who were randomly allocated to either the PBL programme or the traditional

programme. The results of the comparison between the non-randomised groups were not included as this data was produced from what is in effect a cohort study and therefore does not meet the review inclusion criteria.

The study carried out by Verhoeven and colleagues (1998) compares the Maastricht medical school programme that uses PBL, with medical school programmes at other Dutch Universities that do not use PBL. Entry to medical school in the Netherlands is by weighted lottery i.e. students that meet the entry criteria are allocated to each school centrally. The students and the school have no choice. On this basis the authors argue that it is reasonable to assume that the students at all medical schools will be equally able on entry to the programme and any effects seen thereafter can reasonably be attributed to the medical education programmes. Thus the study design was designated as quasi randomised (CCT). The instrument used in the comparisons between Maastricht and other medical schools was the Maastricht Progress Test. This multiple choice test is used for assessment of all students in the Maastricht programme but only as part of the research study in the comparator medical schools. Therefore follow-up rates in the comparator schools tend to be lower. For the included study (Verhoeven et al. 1998) the only effects reported are those where the follow-up (response) rate is greater than 60%.

There is considerable heterogeneity in the outcomes reported and instrumentation used. There is no consensus on the either the outcomes or methods of measurement that should be used to assess the effects of PBL. An important aspect of PBL philosophy is the recognition of the fact that assessment has a major impact on learning. Although all advocates of PBL share this premise its consequences are interpreted differently. Some writers suggest that both the response format and the content of the test must be appropriate to PBL (Marks-Maran & Gail Thomas 2000). Others argue that response format is of less consequence than content and test-design (Norman 1991).

The National Board Medical Exam (NBME) is the state licensing exam for doctors in the USA and all three parts must be passed to qualify as a doctor. The NBME is thus an important outcome measure for students and medical colleges. The NBME uses a multiple-choice format and is composed of a number of sub-scales on different content areas of the medical curriculum. It is a standardised instrument set nationally so it is possible to compare results between different medical colleges and programmes within them. However the NBME exam developed and changed over time (*personal communication Charles Engel 2003*) suggesting that comparison of results obtained at different times may be problematic. The results from the Harvard study (Moore et al 1994) are presented as standardised scores to a mean of zero and standard deviation of 1. The Maastricht Progress Test (MPT) also uses a multiple-choice format and students are repeatedly tested throughout their medical school career. The MPT scores are not used as part of the state licensing process (*personal communication Charles Engel 2003*).

The assessment instrument used by Doucet and colleagues (1998) consisted of a 50 item true/ false questionnaire specifically designed for the content of the programme. The assessment instrument used in the study by Premi and colleagues (1994) was a 40 item multiple choice questionnaire designed by a neurologist with an interest in headache management. The multiple choice instruments used in the studies by Tan et al (1986) and Antepohl & Herzig (1999) are not discussed in any detail. As noted above some commentators have argued that the multiple choice format is not suitable for the assessment in PBL programmes.

The study by Antepohl & Herzig (1999) also used the written test as an assessment format. With their emphasis on self selection of topic, self-directed information searching and presentation of data in a clear focussed manner, written assignments are viewed as a relevant evaluation method within the PBL

approach (Rideout & Carpio 2001). However, problems with reliability of these methods, particularly when the sample of items is small, are well known (Van Der Vleuten 1996). None of the studies included provide any further detail of the assessment. In the case of the written assignments used in the study by Antepohl & Herzig reliability is not established. The study by Doucet and colleagues (1998) also included a written assessment called a 'Key Feature Problem' examination designed to assess clinical reasoning skills. However this assessment was not blinded and nor was reliability estimated and the response rate was less than 80% thus the data were not extracted.

Table seven reports data from studies that attempted to assess 'improvement in practice'. The included studies used a variety of approaches to the assessment. Only effects that met the review inclusion criteria are reported. Other effects were excluded for a combination of low response rates, lack of blinding and/or reliability. The ATSIM instrument used in the Harvard study is described as a measure of students attitudes towards social issues in medicine and the study authors hypothesize that the PBL students should have a higher rating score on these scales (Block et al. 1993). In the study by Lewis and colleagues (Lewis & Tamblyn 1987) a nursing process assessment was used as one of the outcome measures. This is described as a standardized evaluation form completed by students whilst in clinical area and students are graded on the different aspects of the nursing process listed in the table. In this study markers were not blind and neither was inter-rater reliability established.

In the study by Grol and colleagues (1989) PBL acts a control comparator to an intervention involving systematic training and peer review of consultation skills. In the table the effect is reported to fit with the question for this Systematic Review and therefore PBL group results are listed in the intervention columns. The effect measure included from this study is the proportion of obligatory actions completed in a medical consultation. This figure is based on the number of actions taken by the 'students' in a consultation with a real patient against a protocol of 'required' actions for a particular medical condition that was developed by a panel of 'experts'. The students were videoed in real consultations pre and post practice and the grading completed by the University 'experts from the videos. The analysis is based on rate of compulsory actions per consultation in each group not for each student. The number of consultations analysed was 631 pre and 624 post intervention. However this only about half the total number of consultations that there should have been if all the students completed the 20 pre and post intervention videos. Clearly not all video consultations for all students were included in the analysis.

Results - Included study synthesis heterogeneity & effects

The results reported in each study are reported in the same units as in the original publication. Only those effects that meet the inclusion criteria are reported in tables five to nine. Where sufficient data were reported standardised effect sizes (*d*) were calculated. The effect size is the standardised mean difference between the two groups and thus provides an estimate of the size of any difference. One particular advantage of this approach is that we do not need to be familiar with the scale used by the researcher (e.g. NBME) in order to interpret the result (Coe 2002). There are a number of approaches that can be used to calculate effect sizes. A small sample of the effect sizes were calculated using both the pooled standard deviation and the control group standard deviation as the denominator. Even where the difference between the standard deviations in the intervention and control groups was comparatively large (as in the case of Mennin NBME I) the use of the different methods made little difference to the resulting effect size. The Meta-Stat software used for the meta-analysis used the control group standard deviation as the denominator (Rudner et al. 2002a) therefore this method was used for calculating effect size throughout.

Meta-Analysis procedure

Meta-analysis is a formal statistical analysis of the data from the various subgroups of the studies included to get an overall estimate of the effectiveness of the intervention (Altman 1991). Typically the pooled effect estimate represents a weighted average of all studies included in the meta-analysis with a greater weighting being given to larger studies and less weight to smaller studies. The pooling can be carried out using either a fixed effects or random effects model. The fixed effects model estimates the treatment effect as if there were a single 'true' value underlying all the study results. A random effects model assumes that there is no single underlying value of effectiveness but a distribution of values depending on known and unknown characteristics of studies. There is disagreement about which statistical model is appropriate and in practice both statistical models produce similar results (Khan et al. 2001). There is similar controversy over whether the pooled variance or the control group variance should be used in the calculation of effect sizes (Rudner et al 2002a). In the software used (Meta Stat) the effect size is based on the pooled variance of the control group and the effect size used is not corrected for sampling or measurement error (Rudner et al. 2002b).

What evidence do the included effects provide to answer the review questions?

The 'goals of PBL' used as framework for the review question was intended to act as a heuristic device to aid the discovery of patterns of results. However, the majority of the studies included did not report effects in this way. For the purpose of analysis effects were categorised into broader, less specific categories of 'accumulation of knowledge', 'improving practice', 'improving approaches to learning' and 'improving student satisfaction'. Any such categorisation is open to challenge on conceptual, methodological and practical grounds. However it is not suggested that the categories used here are definitive but rather should be viewed as a reasonable pragmatic response to how the individual studies are presented.

Does the use of PBL result in greater 'accumulation of knowledge?

There is an argument that if PBL does claim to improve 'the accumulation of knowledge' it is the kind of knowledge that is manifest in contextualised practice which is not simply the result of the accumulation of factual information but rather a transformation of the individual (Hmelo & Evenson 2000). This is the basis for some of the critiques of the use of the Multiple-Choice Question format to assess PBL (Van Der Vleuten 1996). However, the majority of effects reported in the included studies use this format. Some of the reported effects use 'written tests' which presumably used the free text format that some authors suggest is congruent with the aims of PBL (Marks-Maran & Gail Thomas 2000). Whatever the inadequacies of these means of assessment for evaluating the impact achieved by PBL these methods are used by authors whose reported aim is to do just that. The implication of the use of different formats can be explored through meta-analysis (see below). However, it would seem reasonable to assume that these tests aim to assess at least elements of performance abilities in critical reasoning, problem solving, creativity and decision making which can be grouped under the heading the 'accumulation of knowledge'.

The effects considered under this heading are reported in tables five and six. The data in the tables are self-explanatory with possible exception of the NMBE I effects reported by Moore and colleagues (1994). The data reported in this study are the mean scores standardised to a value of 0 and standard deviations

standardised to a value of 1 (Moore et al 1994). Therefore the figure for the difference in the mean score between the intervention and control groups (column 1-2) represents a standard deviated effect size (*d*).

Figure one shows the effect sizes for 'knowledge' effects with their respective 95% confidence intervals for the 'Total scores' in ascending order. The 'study id' is equivalent to that given in tables four and five. It is not possible to calculate standard deviated effect sizes for the effects reported in the study by Farquar (1986) as insufficient data are reported by the authors. The general impression of the results of these studies gained from the tables is that they indicate that outcomes for students in the PBL groups were less favourable. Of the 39 effects reported in tables four and five, 16 favour the intervention group and 23 the control group. As figure one indicates for effects based on 'Total' scores the balance is more even with seven results favouring the intervention group and nine the control.

Figure 1: Effect sizes with 95% confidence intervals for 'accumulation of knowledge' 'Total' effects only

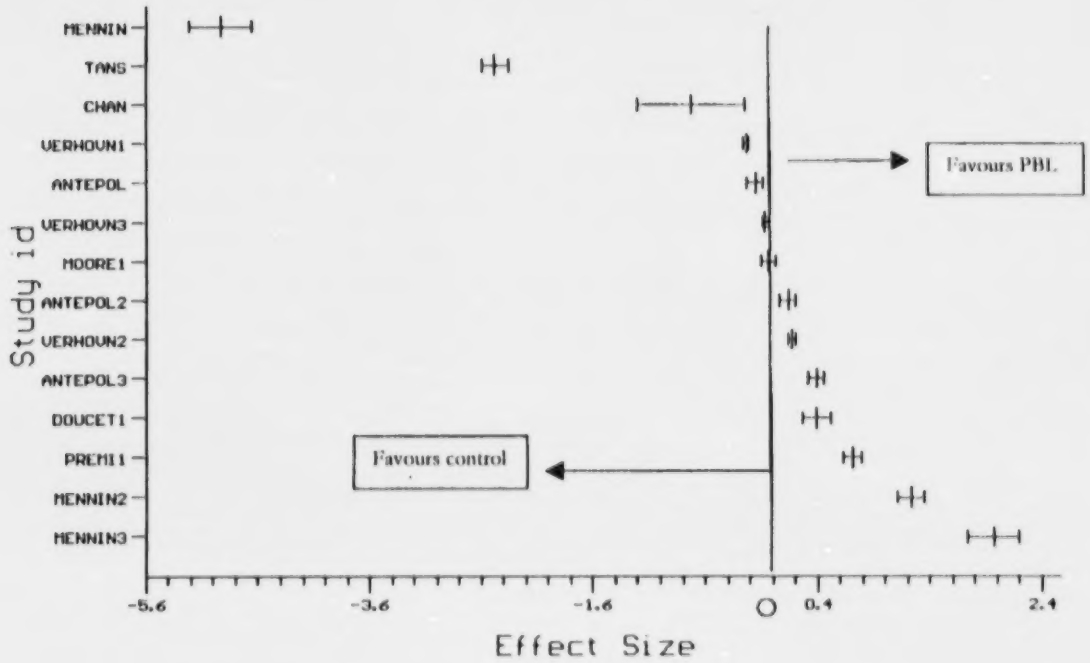


Figure one also illustrates just how different the results reported for the NMBE I assessment by Mennin and colleagues (1999) are from the rest of the results. This result would seem to be clearly separated from the other studies. Values this separated raise suspicion of error in either reporting or calculation (Rudner et al 2002a). The data extraction and calculation used for this report were re checked but found to be satisfactory but the possibility of error in the reporting of the primary study has not been excluded.

A meta-analysis example

Table 10 provides results of an example meta-analysis for the effects included in the category 'accumulation of knowledge'. The analysis is based on the data from all the studies in tables five and six with the exception of the data from the study by Farquar and colleagues (1999). The example meta-

analysis contravenes one of the principles of meta-analysis that all effects included in the meta-analysis should be 'independent' i.e. not from the same subjects (Hedges 2003), however it is not clear that this principle extends to the same subjects at different times as in this example. The estimate of mean effect size (d = -0.3) in this example is based on a fixed effects model.

Table 10: Meta-analysis: Weighted mean effect sizes for category 'accumulation of knowledge'

Moderator	Grouping	Mean Effect size	St. Dev.	N	95% C.I
	Outcome-Knowledge	-0.3	15.81	1904	-1.0 to 0.4
Study Design	Experiment	-0.4	16.47	1719	-1.1 to 0.4
	Quasi – experiment	0.6	6.99	185	-0.4 to 1.6
Randomisation	Random	-0.8	24.63	757	-2.6 to 1.0
	Non – random	0.1	3.77	1174	-0.1 to 0.3
Assessment format	MCQ	-0.3	16.79	1676	-1.1 to 0.5
	Written assessment	0.3	3.4	228	-0.1 to 0.74
Qualification stage of student	Pre – qualification	-0.4	16.56	1700	-1.1 to 0.7
	Post – qualification	0.5	6.7	204	-0.4 to 1.4

The subgroup analysis carried out can be viewed as a form of sensitivity analysis. Not included in the table was a sensitivity analysis of the contribution to the mean effect made by the NBME I effect reported by Mennin and colleagues (Mennin et al 1993). The analysis suggests that this particular result does affect the estimates of the mean effect size as with this result excluded the mean effect sizes for the category 'accumulation of knowledge' and the subgroups moves very close to 0. However, the even with this result excluded the confidence interval still does not exclude the possibility of a large negative effect size (d= -1.0).

Does the use of PBL in 'greater improvements in practice'?

Only three of the studies report data that can be interpreted as measures of 'improvements in practice' these are shown in table six. Only the study by Moore and colleagues (1994) provides sufficient data for the calculation of effect sizes. In this study the outcome refers to attitudes toward practice and for the three effects that met the review inclusion criteria the effect sizes are favour PBL. Of the seven effects reported in the study by Lewis and colleagues (1987) two favour the PBL group. Of the nine effects reported by Grol and colleagues (1989) only one favours the intervention (PBL group).

Does PBL result in better 'approaches to study'?

Two of the included studies assessed changes in student learning styles these are reported in table seven. The Harvard study tested students learning style preferences using the Preferred Learning Style Index. The authors state that that PBL should cultivate the 'Discovery style' of learning (Moore et al 1994). The effect sizes on the discovery scale (d=1.0) and on the receptive scale (d=0.5) favour PBL. The other study in this table compares students learning styles using an instrument called the Short Inventory of Approaches to Studying. On this instrument a positive outcome is a decrease in 'reproducing' and increase in 'meaning' and 'versatility' (Coles 1985). Again the results favour the PBL group with an effect size of d= -1.1 on the reproducing scale, d= 0.5 and d= 0.4 on the meaning and versatility scales respectively. It is interesting to note that on the 'desirable' scales in each study ('Discovery', 'Meaning', and 'Versatility') the mean scores were worse after the intervention in both the PBL and control groups. The decline in the

mean scores was larger in the control groups than in the intervention groups (+1.96, +1.5, +1.8 respectively) suggesting the PBL may be acting to moderate other negative influences on approaches to learning.

Does PBL result in 'greater student satisfaction'?

Whilst student satisfaction was reported in many of the studies considered in the review only in the case of the study by Moore and colleagues (1994) did the effects on this outcome meet the study inclusion criteria. The effects reported in this study are given in table eight. With the exception of the scales 'Task orientation' and 'Clarity' the effect sizes favour the PBL group.

This was a pilot Systematic Review and meta-analysis the aims of which were to demonstrate the need for and the potential benefit of conducting research of this kind on the effectiveness of PBL and to allow the review group to develop and test the procedures for doing so. As such the results, particularly with regard to the meta-analysis should not be interpreted as if they were based on a full systematic review. There was no attempt to systematically identify primary studies from the literature and therefore the included studies represent a bias selection of the existing literature.

Does the level of 'safe' knowledge identified suggest that a full systematic review is required ?

Of the 91 citations identified as providing empirical evidence about the effectiveness of PBL in the five sample reviews only 15 met the review inclusion criteria. Four out of the 30 citations identified from the review by Vernon and colleagues (1993), 4/42 from the Albanese & Mitchell (1993) review, 2/31 in the Berkson (1993) review, 10/38 in the review by Van den Bosche and colleagues (2000) and 4/5 in the review by Smits and colleagues (2002a)[3]. With the exception of the reviews by Van den Bosche et al (2000) and Smits et al (2002a) the reviews are all now more than 10 years old. All of the reviews focus on healthcare education and predominantly on medical education.

One of the most striking features of the studies that met the inclusion criteria and from which data could be extracted was the almost exclusive focus on reporting outcomes from tests or assignments that arguably are measuring 'the accumulation of knowledge' and in particularly the use of multiple choice formats. In part this reflects the fact that the review restricts inclusion to high quality studies. But the lack of reported high quality effects on other outcomes and/or using different approaches to assessment that are argued to be more congruent with PBL reflects the methodological and practical difficulties associated with these approaches. It is the case that some of the larger studies at Harvard and New Mexico in particular had attempted to assess other outcomes, but in the case of the New Mexico studies at least, these effects did not meet the study inclusion criteria usually because of inadequate follow-up rates. Should such a situation persist when a full literature search had been completed, a review team might wish to consider 'relaxing' inclusion quality criteria. The impact of such a strategy could be empirically explored by weighting studies by response rate for example. Lowering the quality threshold for the inclusion of studies may however solve one problem at the expense of another. The limited sensitivity analysis carried out as part of the meta-analysis supports existing evidence that suggests that the quality and size of a study influences effect size (Moore & McQuay 2002) and that non randomised studies tend to report larger estimates of benefit (Reeves et al. 1998).

The aggregation of effects into the category 'accumulation of knowledge' provided a sufficient number of effects to allow an example meta-analysis to be completed. In the case of this outcome the effects reported include results that favour PBL and results that favour the control. Researchers in education and other fields continue to debate the practical significance of an effect size. A general recommended guideline across disciplines is that d=0.2 (small effect), d=0.50 (moderate effect), and d=0.80 (large effect) (Cohen 1988). It has been argued that an effect size of d=2.0 should be required where wholesale curriculum and organisational change is implied.(Bloom 1984). In the case of PBL Colliver argued that d=1.0 should be the minimum size of effect required to justify the considerable change that PBL requires (Colliver 2000). Of the 48 effects extracted from the included studies for which standard deviated effect sizes could be calculated 17 were within a range d=0 to 0.3. Given the problems of measurement error,

[3] Some citations appear in more than one review

non response and the fact that the instruments are measuring latent variables these differences could be reasonably be accounted for by such study artefacts (Gorard et al. 2002). Six of the effects did meet or exceed Colliver's threshold. However of these the two largest effect sizes favoured the control group.

It is worth noting however what some of these reported effect sizes might mean in practice should the true population effect size be anywhere near the effect sizes obtained in these individual studies. Effect sizes can be directly converted into statements about the overlap between two samples in terms of comparison of percentiles (Coe 2002). For example the effect size of d=2.0 obtained for the NMBE III in the study by Moore and colleagues (1994) means that 98% of the control group would be below the average person in the PBL group. Another implication of this interpretation is that changes in the percentage of people reaching a particular threshold can be translated to and from effect sizes (Coe 2002). If an institution with an average pass rate of 50% implemented a change with an effect size of d= 2.0, assuming other things being equal and assuming the effect rate applied equally across the curriculum the pass rate would rise to 98%. Of course the corollary is that if the effect size were negative i.e. in favour of the control group, implementing a change would have the effect of reducing the pass rate.

Meta-analysis of these results of necessity required the pooling of results from studies that varied in their operationization and measurement of independent and dependent variables and that employed different types of sampling units. To some extent this problem was reduced by the inclusion of only studies that met strict methodological inclusion criteria. The appropriateness of combining data from studies where there is variation in the characteristics of the intervention, control, sample and institutional contexts is a matter of controversy in educational research. It could be argued that the pooling of results of studies which use different assessment formats, different levels of education, different disciplines and quite possibly different forms of PBL is analogous to taking apples and oranges and averaging such measures as their weights, sizes, flavours and shelf lives. The figures arrived at might be meaningless. However, it can also be argued that it is a good thing to mix apples and oranges, particularly if one wants to generalise about fruit, and that studies that are exactly the same in all respects are actually limited in generalisability (Rosenthal & DiMatteo 2001).

To extend this analogy further the intention of this systematic review is to compare fruit with vegetables i.e. PBL with teaching and learning strategies that are not PBL. From this perspective aggregating effects from different kinds of PBL is not necessarily problematic. In these circumstances if an effect is detected it more likely to be independent of the specifics of context and variation in design (Hedges 2003). Such a result is not particularly useful where the precision of the estimate does not allow the exclusion of either positive or negative effects. Meta-analysis is valuable here as it increases the precision of the estimate. There does not seem to be a general answer to this issue that can be applied to all meta-analysis. The effect of any differences can only be established through empirical investigation in a specific field. A full systematic review of the effectiveness of PBL will provide more evidence about different kinds of PBL in different settings thus allowing more systematic exploration of the different facets of course organisation through meta-analysis.

Lessons for a Systematic Review

The results of the Pilot review suggest that on the basis of existing reviews there are considerable gaps in our knowledge about the conditions under which PBL can be expected to produce more beneficial outcomes than other strategies of teaching and learning. Moreover existing reviews and specifically the better quality studies within them appear to provide evidence largely about the effects of PBL on students

'accumulation of knowledge'. As such the reviews provide very little evidence regarding the wider, and some would argue more important, claims made for PBL. However this does not mean that evidence to partly complete these gaps does not exist. A Systematic Review that crosses subject and disciplinary boundaries and includes studies only of high methodological quality could go some way to addressing these gaps.

The outcome of the pilot review indicates that the Systematic Review approach of the Cochrane Effective Practice and Organisation of Care Group is applicable to broader Systematic Reviews of education with only minor modification. The pilot suggest that setting inclusion criteria on the basis of the nature of the intervention will continue to be problematic until researchers and publishers start reporting more detail about the educational interventions that are being compared. This could be achieved by Journal Editors and authors adopting agreed standards for the reporting of studies such as the CONSORT statement (Moher et al. 2001). However this will not in itself be sufficient as there appear to additional conceptual and practical problems in relation to the identification of PBL interventions.

To return to the fruit and vegetable analogy there seem to be two problematic issues that are intimately related. Firstly distinguishing between fruit and vegetables, secondly distinguishing between different types of fruit. It might seem premature on the basis of this review to generalise to the wider PBL literature. But our collective experience leads us to suggest that the apparent distinctions between different forms of PBL offered by the use of such terms as 'Authentic', 'hybrid', 'Enquiry based' and 'Inquiry based' are conceptually and practically blurred and in any case not agreed by the different PBL communities. Similarly the view non that non PBL programmes = 'didactic lectures' (= vegetable) this would seem less relevant today particularly outside of medical education. The review raises questions about the usefulness of distinctions such 'authentic' and 'hybrid' as the sole form of distinction given the very different educational contexts in which PBL operates. This is not a problem of Systematic Reviews per-se but rather is an issue for all those interested in PBL. An approach such as that exemplified in Barrows (2000) 'Criteria For analysing a problem based learning curriculum' offers a potential framework of classifying different educational interventions the differential impacts of which could explored empirically through meta-analysis.

The pilot review was an important process for the review team members to go through together as it highlighted different interpretations of terminology and criteria that only a thorough testing of the process could have detected. It has also provided food for thought about a number of methodological and practical issues. Interpretation of how a study was actually carried out may vary between reviewers even where the protocol guidelines appear to be clear. It is suggested that this is due to a combination of factors including the different first languages of the reviewers (to the language that papers are written in and to each other), variety of research terminology, and the real uncertainty that exists in the field. There does not appear to be a general solution to this but a specific development to help with this problem will be the incorporation of a study quality scoring system into the full review process. This will allow the effects of different methodological deficiencies in the included studies to be estimated. However, the tension between the desire to include as many studies as possible (and thus to possibly increase the external generalizability of any review) and the need to keep any review within manageable boundaries is likely to remain a constant tension in reviews of this type.

The practical issues chiefly concern resources. The review co-ordinator spent a considerable amount of time working on the Systematic Review as part of a larger Project on the Effectiveness of Problem Based Learning. The other reviewers also spent not inconsequential amounts of time reviewing papers, and providing feedback and commentary regarding the analysis of the data and report writing. This provides

another practical justification for identifying clear methodological criteria by which to limit the studies that will be included. However, if as the pilot study seems to suggest, additional time will have to be spent reading other reports and contacting study authors to obtain the necessary descriptive information this will add considerably to the time required to complete the review. Although many Journals are now making their contents available on –line it will take many years before the entire back catalogues of journals are available in digitised form. Libraries will also therefore require funding to pay for the cost of obtaining papers that are not available in the host institution or on-line. A full review is likely to require a significant funding investment that will increase inversely with the time allowed for completion. This will include funding for a review coordinator, for expert input from an information professional, time for reviewers, and library costs.

Table 4: Curriculum design and context included studies.

Author	Country	Profession	Subject	Level	Time span	Contact time I	Contact time C	Tutor background	PBL description	Control teaching method	Student/ teacher ratio I	Student/ teacher ratio C	Tutorial process	Method of assessment
Moore et al	USA	1	Medicine	2	2 yrs	?	?	9	2	1	Small group	?	Analyse case material set learning goals	NBME
Mennin et al	USA	1	Medicine	2	2 yrs	5x half day (2-3 PBL sessions)	4 day week	3	2	3	1/5	?	Discuss case materials (McMaster model)	NBME
Verhoeven	Netherlands	1	Medicine	2	6 yrs	?	?	9	2	9	?	?	Maastrict 7 step	MCQ
Tans	Netherlands	4	Physiology	2	7 wks	2hr x 1per week		1	2	1	8/1	15/1	Barrows/Schmidt	MCQ
Antepohl	Germany	1	Basic Pharmacology	2	Semester	3 hours x 1 per week		1	2	1	9/1	20/1	Modified Maastrict	MCQ & SEQ
Coles	England	1	Medicine	2	1 yr	?	?	9	9	9	?	?	?	?
Doucet	Canada	1	Headache diag. & mgt	4	3 mth	3x2 hr	3x 2hr	9	1	3	7/?	13/?	Discuss case materials	MCQ
Farquhar	USA	1	Bio, Clin, Behav sciences	2	Pre-clin. Years	PBL = 4, total = 12 hr week	Tot 25 hr/ wk	1	2	1	12/2	40/?	Case based problem solving	NBME
Chan	Canada	1	Depression in the Elderly	5	2 mth	On –line asynchronous	On-line asynchronous	1	1	4	11/3*	No teacher	Case discussion with tutor e-mailing to present/elaborate problem or redirect groups	MCQ
Grol	Netherlands	1	Consultation skills	4	1 yr	1 day / week		9	9	4	?	?	?	Peer observation
Lewis	Canada	2	Care of acutely ill adult	2	12 wks	4hr/ week	4hr/ week	1	2	3	12/1	20/1	Discuss case material	MCQ
Premi	Canada	1	CME	4	8 Mths	1.5hr x2 month	na	3	1	na	9/1	na	McMaster PBL	None

Table 5: Reported results for experimental studies in category 'accumulation of knowledge'

Author	Design	Random	Comparison	Sample size Int. Start/Effect	Sample size Control Start/Effect	Similarity of control	Assessment Blinding	Contamination	Baseline	Reliability	Outcome measure	1- Int. group (s.dev)	2- Cont. group (s.dev)	1-2	Effect size *d*
Mennin 1	RCT	9	2	?/85	?/34	1	1	1	2	1	NBME I	455 (8.5)	521 (13.4)	-66*	-4.9
Mennin 2				?/67	?27						NBME II	485 (15.9)	472 (10.2)	13	1.3
Mennin 3				?/38	?/19						NBME III	551 (20.7)	521 (14.8)	30	2.0
				?/85	?/34						NBME I pass %	82	97	-15%	
Moore 1	RCT	9	2	62/60	63/61	1	1	1	1	1	NBME I (total)	0.06 (1.09)	0.07 (1.01)	-0.01	
											NBME I subscales: Anatomy	-0.13 (1.0)	0.13 (1.01)	-0.26	
											Behav.Sci	0.37 (0.94)	-0.09 (1.01)	+0.46 *	
											Biochem	-0.04 (1.13)	0.09 (0.94)	-0.13	=effect size
											Microbi	0.16 (1.08)	-0.16 (0.99)	+0.32	
											Pathology	0.13 (1.02)	0.1(1.04)	+0.03	
											Pharm	-0.07 (1.11)	0.01 (1.09)	-0.08	
											Physiology	0.06 (1.06)	0.2 (0.88)	-0.14	
Tans	RCT	1	2	77/74	45/45	1	1	9	3	1	60 item MCQ	27.7 (4.48)	39.2 (4.68)	-5.22*	-2.5
Chan	RCT	9	4	11/8	12/11	2	1	2	1	1	MCQ	64.3 (14.2)	69.3(6.9)	-5	-0.7
											Pre-post Change	-2.3	+3.5	-5.8	0.7#
Antepohl 1	CCT	na	2	63/57	60/57	1	1	1	1	1	MCQ	11.6 (2.7)	12 (3.2)	-0.4	-0.1
Antepohl 2											Written test	22.8 (6.3)	21.8 (6)	+1	0.2
Antepohl 3											Short Essay	11.2 (4.2)	9.8 (3.3)	+1.4	0.4
Verhoeven 1	CCT	na	2	?/190	?/124	2	1	1	3	1	MPT total: Year 1	9.6 (4.06)	10.7 (5.39)	-1.	-0.2
Verhoeven 2				?/146	?/104				MPT total		Year2	20.7 (4.94)	19.3 (6.68)	1.4*	-0.2
Verhoeven 3				?/144	?/140						Year5	38.5 (8.53)	38.8 (9.16)	-0.3	-0.03
				?/190	?/124						MPT basic sciences: Year 1	8.4 (4.77)	10.7 (6.78)	-2.3	-0.03
				?/146	?/104						Year2	19.1 (8.27)	16.2 (8.59)	2.9	0.3
				?/144	?/140						Year5	30.1 (11.0)	31.2 (11.7)	-1.1	-0.1
				?/190	?/124						MPT Clinical sciences: Year 1	8.2 (4.74)	9.2 (6.65)	-1	-0.2
				?/146	?/104						Year2	19.3 (6.94)	20.5 (8.13)	-1.2	-0.1
				?/144	?/140						Year5	44.6 (9.71)	45.3 (10.4)	-0.7	-0.1
				?/190	?/124						MPT Social Sciences: Year 1	14.9 (9.74)	14 (10.3)	0.9	0.1
				?/146	?/104						Year2	26.9 (11.2)	21.8 (10.4)	5.1*	0.5
				?/144	?/140						Year5	41.4 (13)	39.5 (12.1)	1.9	0.2

*statistically significant

Table 6: Reported results for studies using Quasi-experimental designs in category 'accumulation of knowledge'

Author	Design	Random	Comparison	Sample size Int. Start/Effect	Sample size Control Start/Effect	Similarity of control	Assessment Blinding	Contamination	Baseline	Reliability	Outcome measure	1- Int. group (s.dev)	2- Cont. group (s.dev)	1-2	Effect size *d*
Farquhar	CBA	na	2	40/40	40/40	1	1	1	1		NBME I Total:			5.9	
											NBME I subscales: Anatomy			-2.6	
											NBME I subscales: Physiology			-4	
											NBME I subscales: Biochemistry			-35.5	
											NBME I subscales: Pathology			-20.6	
											NBME I subscales: Microbiology			47.6*	
											NBME I subscales: Pharmacology			21.5	
											NBME I subscales: Behavioral science			-2.9	
Premi 1	CBA	na	1	100/76	52/46	1	1	1	1	1	MCQ test	77 (11.7)	68 (12)	9	0.8
											Pre-post Change	+12.%*	+3%	+9%	-0.8~
Doucet 1	CBA	na	4	37/34	49/29	1	1	1	1	1	Knowledge	33.3 (3.7)	31.4 (4.4)	+1.9	0.4
											Pre-post Change	9.03	7.56	+1.5`#	-0.7~

*statistically significant # Statistically significant using ANCOVA running pre-test as covariant

Table 7: Study design and reported effects in category ' improvement in practice'

Author	Design	Random	Comparison	Sample size: Int. Start/ Effect	Sample size: Control Start/ effect	Similarity of control	Assessment Blinding	Contamination	Baseline	Reliability	Outcome measure	1- Int. group (s.dev)	2- Cont. group (s.dev)	1-2	3-Int. change	4- Cont. change	3-4
Moore et al	RCT	9	2	60/55	61/54	1	1	1	1	9	Dr/Pt relate	7.85 (1.11)	7.35 (1.08)	+0.5*	ATSIM	0.5	Effect size *d*
											Preventive	18.74 (2.68)	18.46 (2.4)	+0.28		0.1	
											Social medicine	17.87 (3.11)	17.67 (2.58)	+0.2		0.1	
Lewis	CBA	na	2	24/22	24/20	1	2	1	1	9	Nursing process skills: Total	77.4	79.8	-2.4			
											Teaching	77.6	76.5	+1.1			
											Counselling	67.1	70.6	-3.5			
											Assessment				6.0	13.4	-7.4*
											Implementing				9.1	13	-3.9
											Charting				14.8	7.9	6.9
											Evaluation				3.5	13	-9.5
Grol	CBA	na	4	31/?	32/?	2	2	2	1	1	Consultation skills: % Total consultation	59.2	65.3	-6.1	1.4	6.5*	-5.1
											% Diagnostic phase	60.9	66	-5.1	0	4.8*	-4.8
											% Therapeutic phase	55.1	63.4	-8.3	4.6	10.5*	-5.9
											% History	61.7	69.4	-7.7	-1.2	4.6*	-5.8
											% Physical examination	65	65.9	-0.9	0.6	3.7	-3.1
											% Education/ information	57	63.4	-16.4	4.5*	9*	-4.5
											% Therapy/medication	74	74.1	-0.1	0.9	9.1*	-8.2
											Median .obligatory Perf.	8.72	9.82	-1.1	0.44	1.24*	-0.8
											Median. Unnecessary perf	0.82	0.9	-0.08	-0.83*	-0.42*	+0.41

*statistically significant

37

Table 8: Study design and reported effects in category 'Approaches to learning'

Author	Design	Random	Comparison	Sample size Int. Start/ Effect	Sample size Control Start/ effect	Similarity of control	Assessment Blinding	Contamination	Baseline	Reliability	Outcome measure	1- Int. group (s.dev)	2- Cont. group (s.dev)	1-2	3-Int. change	4-Cont. change	3-4	Standard deviated effect size *d*	Gain effec size
Moore 2	RCT	9	2	61/61	61/58	1	9	1	1	1	Discovery	31.4 (3.75)	27.2 (4.06)	+4.15	-0.27	-2.23	+1.96	1.0	-0.5
Moore 4											Receptive	16.64 (2.51)	17.98 (2.64)	-1.34	-0.61	-0.58	-0.03	-0.5	0.02
Coles 1	CBA	na	2	?	?	2	1	2	1	1	Reproducing	10.8 (3.1)	14.6 (3.6)	-3.8*	-1*	+1.3*	-2.4	-1.1	0.7
Coles 3											Meaning	15.7 (4.1)	13.7 (3.9)	2*	-0.9	-2.4*	+1.5	0.5	-0.5
Coles 5											Versatility	32.9 (6.3)	30.2 (6.3)	2.7*	-1.5	-3.3*	+1.8	0.4	-0.3

*statistically significant

Table 9: Study design and reported effects in category 'satisfaction with learning environment'

Author	Design	Random	Comparison	Sample size Int. Start/ Effect	Sample size Control Start/ Effect	Similarity of control	Assessment Blinding	Contamination	Baseline	Reliability	Outcome measure	1- Int. group (s.dev)	2- Cont. group (s.dev)	1-2	3-Int. change	4-Cont. change	3-4	Standard deviated Effect size *d*
Moore et al	RCT	9	2	36/26	26/25	1	1	1	1	9	Involve	7.61 (1.73)	6.48 (1.9)	1.13*	-0.86	-1.56	+0.7	0.6
											Peer cohesion	7.43 (1.5)	7.24 (1.54)	0.19	-0.1	-0.34	+0.24	0.1
											Faculty support	6.54 (1.62)	6.04 (1.81)	0.5	-0.38	-0.01	+0.37	0.3
											Autonomy	6.61 (1.45)	5.68 (1.6)	0.93*	-1.58	-1.21	-0.37	0.6
											innovation	7.71 (1.49)	3.6 (2.52)	4.11*	+0.82	-2.36	+3.18	1.6
											Task orientation	5.07 (2.28)	5.76 (1.51)	-0.69	-0.79	-0.47	-0.32	-0.5
											Work pressure	5.68 (2.00)	5.68 (1.73)	0	-0.79	-0.47	-0.32	0
											Clarity	3.46 (1.99)	5 (1.56)	-1.54*	-1.85	-0.31	-1.54	-1.0
											Control	2.96 (1.43)	2.76 (1.42)	0.2	-0.57	-0.66	+0.99	0.1

Rating of learning environment

*statistically significant

Albanese, M. A. & Mitchell, S. 'Problem-based Learning: A Review of Literature on its Outcomes and Implementation Issues', *Academic Medicine*, vol. 68, no. 1, (1993), pp. 52-81.

Altman, D. *Practical statistics for Medical research*, 1 edn, (London, Chapman Hall, 1991)

Antepohl, W. & Herzig, S. 'Problem-based learning versus lecture-based learning in a course of basic pharmacology: a controlled, randomized study', *Med.Educ.*, vol. 33, no. 2, (1999), pp. 106-113.

Armstrong, E., Arky, R., Block, S., Federman, D., Huang, A., McAdle, P., & Moore, G. 'Curriculum design', in *New pathways to Medical Education: Learning to learn at Harvard Medical School*, D. Tosteson, S. Adelstein, & S. Carver, eds. (Cambridge MA, Harvard University Press, 1994) pp. 48-77

Barrows, H. S. 'A Taxonomy of Problem-based Learning Methods', *Medical Education*, vol. 20, (1986), pp. 481-486.

Barrows, H. S. *Problem-Based Learning Applied to Medical Education* (Springfield, Southern Illinois University School of Medicine., 2000)

BEME 'Best Evidence in Medical Education: report of a meeting 3-5 December 1999, London UK', *Medical Teacher*, vol. 22, no. 3, (2000), pp. 242-245.

Berkson, L. 'Problem-based learning: have the expectations been met?', *Academic Medicine*, vol. 68, no. 10, (1993), p. S79-S88.

Block, S. D., Style, C. B., & Moore, G. T. 1993, *Can we teach humanism? A randomized controlled trial evaluating the acquisition of humanistic knowledge, attitudes and skills in the new pathway at Harvard medical school* Unpublished paper.

Bloom, B. S. 'The 2 sigma problem: the search for methods of group instruction as effective as one to one tutoring', *Education Research*, vol. 4, (1984), pp. 4-16.

Boruch, R., Bullock, M., Cheek, D., Cooper, H., Davies, P., McCord, J., Soydan, H., & De Moya, D. 2001, *The Campbell Collaboration: Concept, Status and Plans*, Campbell Collaboration Secretariat, Philadelphia, P-653-36.

Boruch, R. F. & Wortman, P. M. 'Implications of education evaluation for education policy', in *Review of research in education*, Berliner D.C, ed. (Washington DC, American Educational Research Association, 1979) pp. 309-361

Chan, D. H., Leclair, K., & Kaczorowski, J. 'Problem-based small-group learning via the Internet among community family physicians: a randomized controlled trial', *MD Comput.*, vol. 16, no. 3, (1999), pp. 54-58.

Coe, R. 'What is an effect size ?', *Building Research Capacity* no. 4, (2002), pp. 6-8.

Cohen, J. *Statistical Power Analysis for the behavioral sciences*, 2 edn, (Hillsdale , NJ, Erlbaum, 1988)

Coles, C. R. 'Differences between conventional and problem-based curricula in their students approaches to studying', *Medical Education*, vol. 19, (1985), pp. 308-309.

Colliver, J. A. 'Effectiveness of problem-based learning curricula: research and theory', *Acad.Med.*, vol. 75, no. 3, (2000), pp. 259-266.

Cook, T. D. & Campbell, D. T. *Quasi-experimentation: design and analysis issues in field settings* (Chicago, Rand McNally, 1979)

health', *British Medical Journal*, vol. 323, (2001), pp. 294-295.

Davis, D. A. & Thomson M.A, O. A. H. 'Changing Physician Performance: a Systematic Review of Continuing Medical Education Strategies', *Journal of The American Medical Association*, vol. 274, (1995), pp. 700-705.

Dewey, J. *Logic, the theory of inquiry* (New York, Holt, 1938)

Dickersin, K., Scherer, R., & Lefebvre, C. 'Identifying relevant studies for systematic reviews', *British Medical Journal*, vol. 309, (1994), pp. 1286-1291.

Doucet, M. D., Purdy, R. A., Kaufman, D. M., & Langille, D. B. 'Comparison of problem-based learning and lecture format in continuing medical education on headache diagnosis and management', *Med.Educ.*, vol. 32, no. 6, (1998), pp. 590-596.

Egger, M., Juni, P., Bartlett, C., Holenstein, F., & Sterne, J. 'How important are comprehensive literature searched and the assessment of trial quality in systematic reviews? Empirical study', *Health Technology Assessment*, vol. 7, no. 1, (2003),

Egger, M. & Smith, G. 'Bias in location and selection of studies', *British Medical Journal*, vol. 316, (1998), pp. 61-66.

Engel, C. E. 'Not Just a Method but a Way of Learning', in *The challenge of problem based learning*, D. Boud & G. P. Felletti, eds. (London, Kogan Page, 1991) pp. 22-33

English National Board 1994, *Creating Lifelong Learners: Partnerships for Care*, English National Board, London.

EPOC. *Cochrane Effective Practice and Organization of Care Group: the data collection checklist* (Aberdeen, EPOC, 1998)

EPPI Centre. *EPPI-Centre Review Group Manual: Version 1.0*, (London, Social Science Research Unit, 2000).

Farquhar, L. J., Haf, J., & Kotabe, K. 'Effect of two preclinical curricula on NBME Part I examination performance', *J.Med.Educ.*, vol. 61, no. 5, (1986), pp. 368-373.

Glanville, J. & Sowden, A. *Identification of the need for a review*, (York, NHS Centre for Reviews & Dissemination, 2001).

Gorard, S., Prandy, K., & Roberts, K., *An introduction to the simple role of numbers in social science research*, ESRC Teaching & Learning Research Programme Research Capacity Building Network, 2002

Gough, D. & Elbourne, D. 'Systematic research synthesis to inform, policy, practice and democratic debate', *Social Policy and Society*, vol. 1, no. 3, (2002), pp. 225-236.

Grol, R., Mokkink, H., Helsper-Lucas, A., Tielens, V., & Bulte, J. 'Effects of the vocational training of general practice consultation skills and medical performance', *Med.Educ.*, vol. 23, no. 6, (1989), pp. 512-521.

Heale J et al 'A randomized controlled trial assessing impact of problem based versus didactice teaching method in CME', (Washington DC, Association of American Medical Colleges, 1988) pp. 72-77

Hedges, L. 'The basics of Meta-analysis', (Stockholm, 3rd Campbell Collaboration Colloquium, 2003)

Hmelo, C. & Evenson, D. 'Introduction to Problem based learning: Gaining insights on learning interactions through multiple methods of enquiry', in *Problem based learning A research perspective on learning interactions*, D. Evenson & C. Hmelo, eds. (Mahwah, Lawrence Erlbaum, 2000)

Hmelo, C. E. 'Problem-based learning: Effects on the early acquisition of cognitive skill in medicine.', *Journal of the Learning Sciences*, vol. 7, no. 2, (1998), pp. 173-208.

Kaufman, A., Mennin, S., Waterman, R., Duban, S., Hansbarger, C., Silverblatt, H., Obenshain, S. S., Kantrowitz, M., Becker, T., Samet, J., & . 'The New Mexico experiment: educational innovation and institutional change', *Acad.Med.*, vol. 64, no. 6, (1989), pp. 285-294.

Khan, K., ter Riet, G., Glanville, J., Sowden, A., & Kleijen, J. *Undertaking systematic reviews of research on Effectiveness: CRD's guidance for those carrying out systematic reviews*, 2nd edn, (York, NHS CRD, 2001)

Lewis, K. E. & Tamblyn, R. M. 'The problem-based learning approach in baccalaureate nursing education: how effective is it?', *Nurs.Pap.*, vol. 19, no. 2, (1987), pp. 17-26.

Light, R. & Pillemer, D. *Summing up: the science of reviewing research* (Cambridge, Mass, Harvard University Press, 1984)

Marks-Maran, D. & Gail Thomas, B. 'Assessment and evaluation in problem based learning', in *Problem-based learning in Nursing A new model for a new context*, S. Glen & K. Wilkie, eds. (Basingstoke, Macmillan Press, 2000) pp. 127-150

Maudsley, G. 'Do we all mean the same thing by "problem-based learning"?: A review of the concepts and formulation of the ground rules', *Academic Medicine*, vol. 74, no. 2, (1999), pp. 178-185.

Mennin, S. P., Friedman, M., Skipper, B., Kalishman, S., & Snyder, J. 'Performances on the NBME I, II, and III by medical students in the problem-based learning and conventional tracks at the University of New Mexico', *Acad.Med.*, vol. 68, no. 8, (1993), pp. 616-624.

Moher, D., Schulz, K., & Altman, D. 'The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials', *Lancet; :* vol. 357, (2001), pp. 1191-1194.

Moore, A. & McQuay, H. 'Mindstretcher 1- Quality and size', *Bandolier*, vol. 9, no. 3, (2002), pp. 3-4.

Moore, G. T., Block, S. D., Style, C. B., & Mitchell, R. 'The influence of the New Pathway curriculum on Harvard medical students', *Acad.Med.*, vol. 69, no. 12, (1994), pp. 983-989.

Muller, S. 'Physicians for the 21st century : Report of the Project Panel of the General Professional Education of the Physician and College Preparation for Medicine', *Journal of Medical Education*, vol. 59 Part 2, (1984),

Norman, G. 'What should be assessed', in *The challenge of problem based learning*, D. Boud & G. Feletti, eds. (London, Kogan Page, 1991) pp. 254-259

Norman, G. & Schmidt, H. 'The psychological basis of Problem-based learning: A review of the evidence', *Academic Medicine*, vol. 67, no. 9, (1992), pp. 557-564.

Oakley, A. 'Research evidence, knowledge management and educational practice: early lessons from a systematic approach', *London Review of Education*, vol. 1, no. 1, (2003), pp. 21-33.

Petticrew, M. 'Systematic reviews from astronomy to zoology: myths and misconceptions', *British Medical Journal*, vol. 322, (2001), pp. 98-101.

Premi, J., Shannon, S., Hartwick, K., Lamb, S., Wakefield, J., & Williams, J. 'Practice-based small-group CME', *Acad.Med.*, vol. 69, no. 10, (1994), pp. 800-802.

Reeves, B., Maclehose, R., Harvey, I., Sheldon, T., Russell, I., & Black, A. 'Comparisons of effect sizes derived from randomised and non-randomised studies', in *Health Services Research Methods: A guide to best practice*, N. Black et al., eds. (London, BMJ Books, 1998) pp. 73-85

Rideout, E. & Carpio, B. 'The problem based learning model of nursing education', in *Transforming Nursing Education through problem-based learning*, Rideout E, ed. (Sudbury, Jones & Bartlett, 2001) pp. 21-45

Rosenthal, R. & DiMatteo, M. 'Meta-analysis: Recent developments in quantitative methods for literature reviews.', *Annual Review of Psychology*, vol. 52, (2001), pp. 59-82.

Rudner, L., Glass, G., Evartt, D., & Emery, P. 2002a, *Setting up a Meta-analysis*, I.MP Associates, Maryland.

Rudner, L., Glass, G., Evartt, D., & Emery, P. *Users guide to the meta analysis of research studies* (Maryland, ERIC Clearing house on assessment and evaluation, 2002b)

Savin-Baden, M. *Problem-based learning in Higher Education: Untold Stories* (Buckingham, Society for Research in Higher Education and Open University Press, 2000)

Schmidt, H. G. 'Problem -based learning: rationale and description', *Medical Education*, vol. 17, (1983), pp. 11-16.

Schmidt, H. G. 'Foundations of problem based learning; some explanatory notes', *Medical Education*, vol. 27, no. 422, (1993), p. 432.

Schon, D. *Educating the reflective practitioner* (Oxford, Jossey Bass, 1987)

Schwandt, T. 'the interpretive review of educational matters; is there any other kind', *Review of Educational Research*, vol. 68, no. 4, (1998), pp. 409-412.

Shadish, W. & Myers, D. 2002, *Campbell Collaboration Research Design Policy Brief 1*, Campbell Collaboration, Pennsylvania.

Smits, P., Verbeek, J., & De Buisonje, C. 'Problem based learning in continuing medical education: a review of controlled evaluation studies', *British Medical Journal*, vol. 324, (2002a), pp. 153-156.

Tans, R. W., Schmidt, H. G., Schade Hoogeveen, B., & Gijselaers, W. H. 'Sturing van het onderwijsleerproces door middel van problemen: een veldexperiment (guiding the learning process by means of problems: a field experiment)', *Tijdschrift voor onderwijsresearch*, vol. 11, (1986), pp. 35-46.

Tate, R. 'Experimental design', in *Encyclopaedia of education research*, H. Mitzel & W. Rabinowitz, eds. (London, Collier Macmillan, 1982) pp. 553-561

Van den Bossche, P., Gijbels, D., & Dochy, F. 'Does problem based learning educate problem solvers? A meta-analysis on the effects of problem based learning', (VII EDINEB Conference, Newport Beach USA, 2000)

Van Der Vleuten, C. P. M. 'The assessment of professional competence: Developments, research and practical implications', *Advances in Health Sciences Education*, vol. 1, no. 41, (1996), p. 67.

Verhoeven, B., Verwijnen G.M., Scherpbier A.J.J.A., Holdrinet R.S.G., & Oes B. 'An analysis of progress test results of PBL and non-PBL students', *Medical Teacher*, vol. 20, no. 4, (1998), pp. 310-316.

Vernon, D. T. & Blake, R. L. 'Does problem-based learning work? A meta-analysis of evaluative research', *Acad.Med.*, vol. 68, no. 7, (1993), pp. 550-563.

Verwijnen, M., Van Der Vleuten, C., & Imbos, T. A. 'Comparison of an innovative medical school with traditional schools: An analysis in the cognitive domain', in *Innovation in medical education: an evaluation of its present status*, Z. Nooman, H. Schmidt, & E. Ezzat, eds. (New York, Springer Publishing, 1990) pp. 40-49

Vu, N. V. & Galofre, A. 'How medical students learn', *J.Med.Educ.*, vol. 58, no. 8, (1983), pp. 601-610.

Vygotsky, L. *Mind in society: The development of higher psychological processes* (Cambridge, MA, Harvard University Press, 1978)

Walton, H. J. & Matthews, M. B. 'Essentials of Problem Based Learning', *Medical Education*, vol. 23, (1989), pp. 542-558.

Wilkie, K. 'The nature of Problem-based learning', in *Problem based learning in nursing*, S. Glen & K. Wilkie, eds. (Basingstoke, Macmillan Press, 2000) pp. 11-34

Wolf, F. 'Lessons to be learnt from evidence based medicine: practice and promise of evidence based medicine and evidence based education', *Medical Teacher*, vol. 22, no. 3, (2000), pp. 251-255.

Wolf, F. 'Problem-based learning and Meta-analyis: can we see the forest through the trees?', *Academic Medicine*, vol. 68, no. 7, (1993), pp. 542-545.

World Health Organization 1993, *Increasing the relevance of Education for Health professionals: Report of a WHO study group on Problem Solving Education for Health professionals*, WHO, Geneva.

World Bank. 1993, *World Development report 1993: Investing in Health*, Oxford University Press, Oxford.

APPENDIX 1 Coding sheet (for tables 4 to 9)

Reporting review results

This coding is derived from the CRD guidance, EPOC handbook and the 'Campbell Collaboration Research Design Policy Brief'. It reflects the approach to study inclusion adopted by the review group i.e. the EPOC approach. Thus it does not reflect entirely the approach that has been proposed for Campbell reviews in the above document.

Box 1: Research design reporting

i) Kind of design (free text)

ii) Randomisation (RCT design only)
1. Genuinely random and concealed 2. Inadequate randomisation 9.Unknown

iii) Type of comparison condition
1. No Intervention 2.Intervention as usual 3. Untargeted activity 4. Other

iv) Sample Size for Intervention group
Initial sample size for each group/sample size for this effect

v) Sample size for control group
Initial sample size for each group/sample size for this effect

vi) Similarity of control group
1. Another group from the same pool of participants 2.External 3.Archival
4. Other 9.Unknown

vii) Blinding of assessment
1. Assessor blind to status of participant or objective test 2.Assessor aware of status of participant ar.' subjective test 9. Not known

viii) Protection from contamination
1. Intervention and control groups on the same site and/or have the same teacher
2. Intervention and control group on different sites & have different teachers
9.Unknown

ix) Baseline measures
1.Done – Groups appear evenly matched on key characteristics 2. Done - Groups do not appear evenly matched 3. Not done 9. Not known

x) Reliability
1. Done – Objective test or Kappa >0.8 2. Not Done – Not object test or Kappa <0.8
3.Not done 9. Not known

Box 2: Reporting of Description of intervention and control curricula and context of study

Country of study: (free text)

Profession or Discipline of students
1. Medicine 2. Nursing 3. Pharmacists 4. Physiotherapists

Subject of study (free text)

Academic level of course
1. Preregistration (not graduate level) 2. Preregistration
3. Masters or above 4. Certificated CPE 5. Non-certificated CPE 6 Other

Age: Mean (range)

Setting
1. University (including professional schools) 2. Workplace 3. HE College
9. Unknown

Motivation
1. No specific incentive 2. Promotion/ payrise 3. Requirement
9. Unknown

Curriculum time span (free text)

Distribution of contact time (free text)

PBL description
1. PBL sole method teaching & learning for all subjects 2. Subjects taught outside PBL
3. For intervention group PBL is an addition to curriculum 9. Unknown

Control group teaching method
1. Lecture 2. Groupwork 3. Lecture + groupwork 4. Other 9.Unknown

Tutor background
1. Subject expert 2. PBL expert 3PBL & Subject expert 9. Unknown

Student teacher ratio for intervention & control groups (free text)

Tutor Training
1. Received training and development in PBL 2. Not received training & development in PBL
9.Unknown

Tutorial process (free text)
Trigger design
1) Question/ problem given in trigger materials 2. Trigger allows free exploration

Use of patients
1. Real/ simulated patients used 2. Real simulated patients not used
9.Unknown

Method of Student assessment (free text)

Method of student grading
1. Pass/fail 2. Incremental grading system used 9.Unknown

APPENDIX 2: Bibliography of studies considered for the pilot review

Al Haddad, M.K. & Jayawickramarajah, P.T. Problem-based curriculum: outcome evaluation, *Med.Teach.* 13,4, (1991), pp 273 - 279

Albano, M.G., Cavallo, F., Hoogenboom, R., Magni, F., Majoor, G., Manenti, F., Schuwirth, L., Stiegler, I., and van, d., V. An international comparison of knowledge levels of medical students: the Maastricht Progress Test, *Med.Educ.* 30,4, (1996), pp 239 - 245

Anderson, S., Camp, H.G., and Philip, J.R., 'Library Utilization by medical students in a traditional or problem based curriculum', *Teaching & Assessing Clinical Competence,* Eds. Bender, W., Hiemstra, R.J., and Scherpier, A., (Groeningen, Boek Work Publications, 1990) pp 77 - 80

Antepohl, W. & Herzig, S. Problem-based learning versus lecture-based learning in a course of basic pharmacology: a controlled, Randomized study, *Med.Educ.* 33,2, (1999), pp 106 - 113

Baca, E., Mennin, S., Kaufmann, A., and Moore-West, W., 'Comparison between a problem-based community oriented track and a traditional track within one medical school', *Innovation in medical education: an evaluation of its present status,* Eds. Nooman, Z., Schmidt, H., and Ezzat, E., (New York, Springer Publishing, 1990) pp 9 - 26

Barrows, H.S. & Tamblyn, R.M. The portable patient problem pack: a problem-based learning unit, *J.Med.Educ.* 52,12, (1977), pp 1002 - 1004

Benjamin, E.M., Schneider, M.S., and Hinchey, K.T. Implementing practice guidelines for diabetes care using problem-based learning. A prospective controlled trial using firm systems, *Diabetes Care.* 22,10, (1999), pp 1672 - 1678

Bickley, H., Donner, R.S., Walker, A.N., and Tift, J.P. Pathology education in a problem based medical curriculum, *Teaching and learning in Medicine.* 2,1, (1990), pp 38 - 41

Block, S. D, Style, C. B, and Moore, G. T. *Can we teach humanism? A Randomized controlled trial evaluating the acquisition of humanistic knowledge, attitudes and skills in the new pathway at Harvard medical school.* Unpublished Paper 1993

Block, S.D. and Moore, G., 'project evaluation', *New pathways to medical education: Learning to learn at Harvard medical school,* Eds. Tosteson, D.C. and Adelstein Carver, S.T., (Cambridge MA, Harvard University Press, 1994)

Blumberg , P. & Michael, J. Development of self-directed learning behaviours, *Teaching & Learning in Medicine.* 4,1, (1992), pp 3 - 8

Blumberg,P. and Eckenfels,E., *A comparison of student satisfaction with their preclinical environment in a traditional and problem based curriculum, Research in Medical Education 1988; Proceedings of the 27th annual conference,* (Washington DC, Association of American Medical Colleges, 1988)

Boshuizen, H.P. and Schmidt, H., 'Curriculum style and the integration of biomedical and clinical knowledge', *Problem-based learning as an educational strategy,* Eds. Bouhuys, P.A.J., Schmidt, H., and Van Berkel, (Maastricht, Network publications, 1993) pp 33 - 41

Bridgham, R., Solomon, D., and Haf, J. The effect of curriculum era on NBME Part I outcomes in a problem-based versus a traditional curriculum track, *Acad.Med.* 66,9 Suppl, (1991), pp S82 - S84

Chan, D.H., Leclair, K., and Kaczorowski, J. Problem-based small-group learning via the Internet among community family physicians: a Randomized controlled trial, *MD Comput.* 16,3, (1999), pp 54 - 58

Claessen, H.F. & Boshuizen, H.P. Recall of medical information by students and doctors, *Med.Educ.* 19,1, (1985), pp 61 - 67

Clarke, R.M., Peletti, G.T., and Engel, C.E. Student perceptions of the learning environment in a new medical school, *Med.Educ.* 18,5, (1984), pp 321 - 325

Colditz, G.A. The students' view of an innovative undergraduate medical course: the first year at the University of Newcastle, N.S.W, *Med.Educ.* 14,5, (1980), pp 320 - 325

Coles, C.R. Differences between conventional and problem-based curricula in their students' approaches to studying, *Med.Educ.* 19,4, (1985), pp 308 - 309

De Vries, M., Schmidt, H., and De Graf, E., 'Dutch comparisons: Cognitive and motivational effects of PBL on medical students', *New directions for medical education: Problem based learning and community oriented medical education,* Eds. Schmidt, H., Lipkin, J.R., De Vries, M., and De Greep, J., (New York, Springer Verlag, 1989) pp 230 - 238

Des Marchais, J.E., Bureau, M.A., Dumais, B., and Pigeon, G. From traditional to problem-based learning: a case report of complete curriculum reform, *Med.Educ.* 26,3, (1992), pp 190 - 199

Dietrich, A.J., Moore-West, M., Palmateer, D.R., Radebaugh, J., Reed, S., and Clauson, B. Adapting problem-based learning to a traditional curriculum: teaching about prevention, *Fam.Pract.Res.J.* 10,1, (1990), pp 65 - 73

Distlehorst, L.H. & Robbs, R.S. A comparison of problem-based learning and standard curriculum students: Three years of retrospective data., *Teaching & Learning in Medicine.* 10,3, (1998), pp 131 - 137

Donner, R.S. & Bickley, H. Problem-based learning: an assessment of its feasibility and cost, *Hum.Pathol.* 21,9, (1990), pp 881 - 885

Doucet, M.D., Purdy, R.A., Kaufman, D.M., and Langille, D.B. Comparison of problem-based learning and lecture format in continuing medical education on headache diagnosis and management, *Med.Educ.* 32,6, (1998), pp 590 - 596

Drop, M. and Post, G., 'Perceptions and evaluations by graduates and faculty members of the of the Maastricht Problem-Based Curriculum', *Innovation in medical education: an evaluation of its present status,* Eds. Nooman, Z., Schmidt, H., and Ezzat, E., (New York, Springer Publishing, 1990) pp 152 - 164

Eisenstaedt, R.S., Barry, W.E., and Glanz, K. Problem-based learning: cognitive retention and cohort traits of randomly selected participants and decliners, *Acad.Med.* 65,9 Suppl, (1990), pp S11 - S12

Farquhar, L.J., Haf, J., and Kotabe, K. Effect of two preclinical curricula on NBME Part I examination performance, *J.Med.Educ.* 61,5, (1986), pp 368 - 373

Finch, P.M. The effect of problem-based learning on the academic performance of students studying podiatric medicine in Ontario, *Med.Educ.* 33,6, (1999), pp 411 - 417

Goodman, L.J., Brueschke, E.E., Bone, R.C., Rose, W.H., Williams, E.J., and Paul, H.A. An experiment in medical education. A critical analysis using traditional criteria, *JAMA.* 265,18, (1991), pp 2373 - 2376

Gordon, M.J., 'Use of heuristics in diagnostic problem solving', *Medical problem solving: An analysis of clinical reasoning,* Eds. Elstein, A.S., Shulman, L.S., and Sprafka, S.A., (Cambridge MA, Harvard University Press, 1978) pp 252 - 272

Grol, R., Mokkink, H., Helsper-Lucas, A., Tielens, V., and Bulte, J. Effects of the vocational training of general practice consultation skills and medical performance, *Med.Educ.* 23,6, (1989), pp 512 - 521

Hamad, B. Problem-based education in Gezira, Sudan, *Med.Educ.* 19,5, (1985), pp 357 - 363

Heale J et al, *A Randomized controlled trial assessing impact of problem based versus didactic teaching method in CME, Research in Medical Education 1988; Proceedings of the 27th annual conference,* (Washington DC, Association of American Medical Colleges, 1988)

Hmelo, E, Gotterer G, and Bransford J.D. A theory-driven approach to assessing the cognitive effects of PBL, *Instructional Science.* 25,6, (1997), pp 387 - 408

Hmelo, C.E. Problem-based learning: Effects on the early acquisition of cognitive skill in medicine., *Journal of the Learning Sciences.* 7,2, (1998), pp 173 - 208

Imbos, T. and Verwijnen, M., 'Voortagangstoeting aan de medische faculteit Maastricht', *probleemgestuurd onderwijs bidjdragon tot onderwijsreserachdagen,* Ed. Schmidt, H., (? Stichting voor oonderzoek, Van Het Onderwijs, 1982) pp 45 - 56

Imbos, T., Drukker, J., Van Mameren, and Verwijnen, M., 'The growth of knowledge of anatomy in a problem based curriculum', *Tutorials in problem based learning,* Ed. Schmidt, H., (Assen Van Gorcum, 1984) pp 106 - 115

Jones, J.W., Bieber, L.L., Echt, R., Scheifly, V., and Ways, P.O., 'A problem-based curriculum- Ten years of experience.', *Tutorials in problem based learning,* Ed. Schmidt, H., (Assen Van Gorcum, 1984) pp 181 - 198

Kassebaum, D.K., Averbach, R.E., and Fryer, G.E., Jr. Student preference for a case-based vs. lecture instructional format, *J.Dent.Educ.* 55,12, (1991), pp 781 - 784

Kaufman, A., Mennin, S., Waterman, R., Duban, S., Hansbarger, C., Silverblatt, H., Obenshain, S.S., Kantrowitz, M., Becker, T., Samet, J., and . The New Mexico experiment: educational innovation and institutional change, *Acad.Med.* 64,6, (1989), pp 285 - 294

Klass, D. and et al, 'Portability of a multiple station performance based assessment of clinical competence.', *Further developments in assessing clinical competence,* Eds. Hart, I.R. and Hardin, R.M., (Montreal, Can-Healy Publications, 1987) pp 434 - 439

Lewis, K.E. & Tamblyn, R.M. The problem-based learning approach in baccalaureate nursing education: how effective is it?, *Nurs.Pap.* 19,2, (1987), pp 17 - 26

Martenson, D., Eriksson, H., and Ingelman-Sundberg, M. Medical chemistry: evaluation of active and problem-oriented teaching methods, *Med.Educ.* 19,1, (1985), pp 34 - 42

Maxwell, J.A. & Wilkerson, L. A study of non-volunteer faculty in a problem-based curriculum, *Acad.Med.* 65,9 Suppl, (1990), pp S13 - S14

McAuley, R.G. & Woodward, C.W. Faculty perceptions of the McMaster M.D. program, *J.Med.Educ.* 59,10, (1984), pp 842 - 843

Mennin, S. and Martinez Burrola, N., 'Cost of problem based learning', *Implementing problem-based medical education: Lessons from successful innovations,* Ed. Kaufmann, A., (New York, Springer publishing Co., 1985) pp 207 - 222

Mennin, S.P., Friedman, M., Skipper, B., Kalishman, S., and Snyder, J. Performances on the NBME I, II, and III by medical students in the problem-based learning and conventional tracks at the University of New Mexico, *Acad.Med.* 68,8, (1993), pp 616 - 624

Mitchell,R., The development of the cognitive behaviour survey to assess medical student learning, Proceedings of the Annual meeting, (San Francisco, American Educational Research Association, 1992)

Moore-West, M. and O'Donnell, M.J., 'Program evaluation', *Implementing problem-based medical education: Lessons from successful innovations,* Ed. Kaufmann, A., (New York, Springer publishing Co., 1985) pp 180 - 206

Moore, G.T. The effect of compulsory participation of medical students in problem-based learning, *Med.Educ.* 25,2, (1991), pp 140 - 143

Moore, G.T., Block, S.D., Style, C.B., and Mitchell, R. The influence of the New Pathway curriculum on Harvard medical students, *Acad.Med.* 69,12, (1994), pp 983 - 989

Moran, J.A., Kirk, P., and Kopelow, M. Measuring the effectiveness of a pilot continuing medical education program, *Can.Fam.Physician.* 42,(1996), pp 272 - 276

Morgan, H.R. A problem-oriented independent studies programme in basic medical sciences, *Med.Educ.* 11,6, (1977), pp 394 - 398

Neame, R., 'Problem based medical education, the Newcastle approach', *New directions for medical education: Problem based learning and community oriented medical education,* Eds. Schmidt, H., Lipkin, J.R., De Vries, M., and De Greep, J., (New York, Springer Verlag, 1989) pp 230 - 238

Neufeld, V. and Sibley, J., 'Evaluation of health sciences programs: programme and student assessment at McMaster University', *New directions for medical education: Problem based learning and community oriented medical education,* Eds. Schmidt, H., Lipkin, J.R., De Vries, M., and De Greep, J., (New York, Springer Verlag, 1989) pp 165 - 179

Neufeld, V.R., Woodward, C.A., and MacLeod, S.M. The McMaster M.D. program: a case study of renewal in medical education, *Acad.Med.* 64,8, (1989), pp 423 - 432

Newble, D.I. & Gordon, M.I. The learning style of medical students, *Med.Educ.* 19,1, (1985), pp 3 - 8

Newble, D.I. & Clarke, R.M. The approaches to learning of students in a traditional and in an innovative problem-based medical school, *Med.Educ.* 20,4, (1986), pp 267 - 273

Nolte,J., Eller,P., and Ringel,S.P., Shifting toward problem based learning in a medical school neurobiology course, Research in Medical Education 1988; Proceedings of the 27th annual conference, (Washington DC, Association of American Medical Colleges, 1988)

Olson, J.O. The McMaster philosophy: a student's perspective on implementation, *Med.Educ.* 21,4, (1987), pp 293 - 296

Patel, V.L., Groen, G.J., and Norman, G.R. Effects of conventional and problem-based medical curricula on problem solving, *Acad.Med.* 66,7, (1991), pp 380 - 389

Polglase, R., Parish, D.C., and Camp, B. Problem-based advance cardiac life support, *Acad.Emerg.Med.* 3,2, (1996), pp 184 - 187

Post, G. and Drop, M., 'Perceptions of the content of the medical curriculum at the medical faculty in Maastricht: A comparison with traditional curricula in the netherlands', *Innovation in medical education: an evaluation of its present status,* Eds. Nooman, Z., Schmidt, H., and Ezzat, E., (New York, Springer Publishing, 1990) pp 64 - 75

Premi, J., Shannon, S., Hartwick, K., Lamb, S., Wakefield, J. and Williams, J. Practice-based small-group CME, *Acad.Med.* 69,10, (1994), pp 800 - 802

Puett, D. & Braunstein, J.J. The endocrine module: An integrated course for first year medical students combining lecture based and modified problem-based curricula, *Teaching and learning in Medicine.* 3,3, (1991), pp 159 - 165

Rangachari, P.K. Design of a problem-based undergraduate course in pharmacology: implications for the teaching of physiology, *Am.J.Physiol.* 260,6 Pt 3, (1991), pp S14 - S21

Richards, B.F., Ober, K.P., Cariaga-Lo, L., Camp, M.G., Philp, J., McFarlane, M., Rupp, R., and Zaccaro, D.J. Ratings of students' performances in a third-year internal medicine clerkship: a comparison between problem-based and lecture-based curricula, *Acad.Med.* 71,2, (1996), pp 187 - 189

Santos-Gomez, L., Kalishman, S., Rezler, A., Skipper, B., and Mennin, S.P. Residency performance of graduates from a problem-based and a conventional curriculum, *Med.Educ.* 24,4, (1990), pp 366 - 375

Saunders, K., Northup, D., and Mennin, S., 'The library in a problem-based curriculum', *Implementing problem-based medical education: Lessons from successful innovations,* Ed. Kaufmann, A., (New York, Springer publishing Co., 1985) pp 71 - 88

Saunders, N., McIntosh, J., McPherson, J., and Engle, C.A., 'Comparison between University of Newcastle and University of Sydney final year students: knowledge and competence', *Innovation in medical education: an evaluation of its present status,* Eds. Nooman, Z., Schmidt, H., and Ezzat, E., (New York, Springer Publishing, 1990) pp 50 - 54

Schmidt, H.G., Machiels-Bongaerts, M., Hermans, H., ten Cate, T.J., Venekamp, R., and Boshuizen, H.P. The development of diagnostic competence: comparison of a problem-based, an integrated, and a conventional medical curriculum, *Acad.Med.* 71,6, (1996), pp 658 - 664

Schuwirth, L. W. T. *An approach to the assessment of medical problem solving: computerized case-based testing.* (Maastricht, Datawyse, 1998)

Schwartz, R.W., Burgett, J.E., Blue, A.V., Donnelly, M.B., and Sloan, D.A. Problem-based learning and performance based testing: effective alternatives for undergraduate surgical education and assessment of student performance, *Medical Teacher.* 19,(1997), pp 19 - 23

Shin, J.H., Haynes, R.B., and Johnston, M.E. Effect of problem-based, self-directed undergraduate education on life-long learning, *CMAJ.* 148,6, (1993), pp 969 - 976

Son, B. & Van Sickle, R.L. Problem solving instruction and students acquisition, retention and structuring of economics knowledge, *Journal of Research and Development in Education.* 33,2, (2000), pp 95 - 105

Tans, R.W., Schmidt, H.G., Schade Hoogeveen, B., and Gijselaers, W.H. Sturing van het onderwijsleerproces door middel van problemen: een veldexperiment, *Tijdschrift voor onderwijsresearch.* 11,(1986), pp 35 - 46

Tolnai, S. Continuing medical education and career choice among graduates of problem-based and traditional curricula, *Med.Educ.* 25,5, (1991), pp 414 - 420

Van Aalst, V., Chatron, E., Noten, A., and Thoonen, B., 'Problem based learning from a students perspective', *Problem based learning: perspectives from the Maastricht curriculum,* Eds. Van Der Vleuten, C. and Wijnen, W., (Amsterdam, Thesis publishers, 1990) pp 77 - 81

Van Hessen, P.A.W. and Verwijen, G.M., 'Does Problem-based learning provide other knowledge?', *Teaching & Assessing Clinical Competence,* Eds. Bender, W., Hiemstra, R.J., and Scherpier, A., (Groeningen, Boek Work Publications, 1990) pp 446 - 451

Verhoeven B.H., Verwijnen G.M., Scherpbier A.J.J.A., Holdrinet R.S.G., and Oes B. An analysis of progress test results of PBL and non-PBL students, *Medical Teacher.* 20,4, (1998), pp 310 - 316

Vernon, D.T., Campbell, J.D., and Dally, J.C. Problem-based learning in two behavioral sciences courses at the University of Missouri-Columbia, *Acad.Med.* 67,5, (1992), pp 349 - 350

Verwijnen, M., Van Der Vleuten, C., and Imbos, T.A., 'Comparison of an innovative medical school with traditional schools: An analysis in the cognitive domain', *Innovation in medical education: an evaluation of its present status,* Eds. Nooman, Z., Schmidt, H., and Ezzat, E., (New York, Springer Publishing, 1990) pp 40 - 49

Vu, N.V. & Galofre, A. How medical students learn, *J.Med.Educ.* 58,8, (1983), pp 601 - 610

West, D.A., Umland, B.E., and Lucero, S.M., 'Evaluating student performance', *Implementing problem-based medical education: Lessons from successful innovations,* Ed. Kaufmann, A., (New York, Springer publishing Co., 1985) pp 144 - 163

West, D.A. & West, M.M. Problem-based learning of psychopathology in a traditional curriculum using multiple conceptual models, *Med.Educ.* 21,2, (1987), pp 151 - 156

Woodward,C., McAuley,R.G., and Ridge, *Unravelling the meaning of global comparative ratings of interns, Research in Medical Education 1981; Proceedings of the 20th annual conference,* (Washington DC, Association of American Medical Colleges, 1981)

Woodward, C., 'Monitoring an innovation in medical education: the McMaster experience', *Innovation in medical education: an evaluation of its present status,* Eds. Nooman, Z., Schmidt, H., and Ezzat, E., (New York, Springer Publishing, 1990) pp 27 - 39

Woodward, C.A. & Ferrier, B.M. Perspectives of graduates two or five years after graduation from a three-year medical school, *J.Med.Educ.* 57,4, (1982), pp 294 - 302

Woodward, C.A. & Ferrier, B.M. The content of the medical curriculum at McMaster University: graduates' evaluation of their preparation for postgraduate training, *Med.Educ.* 17,1, (1983), pp 54 - 60

Woodward, C.A., Ferrier, B.M., Cohen, M., and Goldsmith, C. Billing patters of general practitioners and family physicians in Ontario: A comparison of graduates of McMaster medical school with graduates of other Ontario medical schools, *Teaching & Learning in Medicine.* 2,2, (1988)

A pilot systematic review and meta-analysis on the effectiveness of Problem Based Learning

Part II: Review Protocol

2 Review questions/objectives

The initial review questions are as follows. Does PBL result in increased participant performance at:

- adapting to and participating in change;
- dealing with problems and making reasoned decisions in unfamiliar situations;
- reasoning critically and creatively;
- adopting a more universal or holistic approach;
- practising empathy, appreciating the other person's point of view;
- collaborating productively in groups or teams;
- Identifying own strengths and weaknesses and undertaking appropriate remediation (self-directed learning)

.....when compared to other non-PBL teaching and learning strategies?

The approach taken to the operationalization and measurement of student performance in these areas is likely to vary between PBL curricula. All reported outcomes will be included in the review. The seven goals identified above will be used as a framework for analysis and synthesis of the findings from individual studies. Steps will be taken to ensure that if possible a secondary review question about whether an 'authentic' PBL curriculum delivers a greater improvement in performance (as defined above) than so called 'hybrid' curricula can be carried out.

3 Methods of review

3.1 Review Process

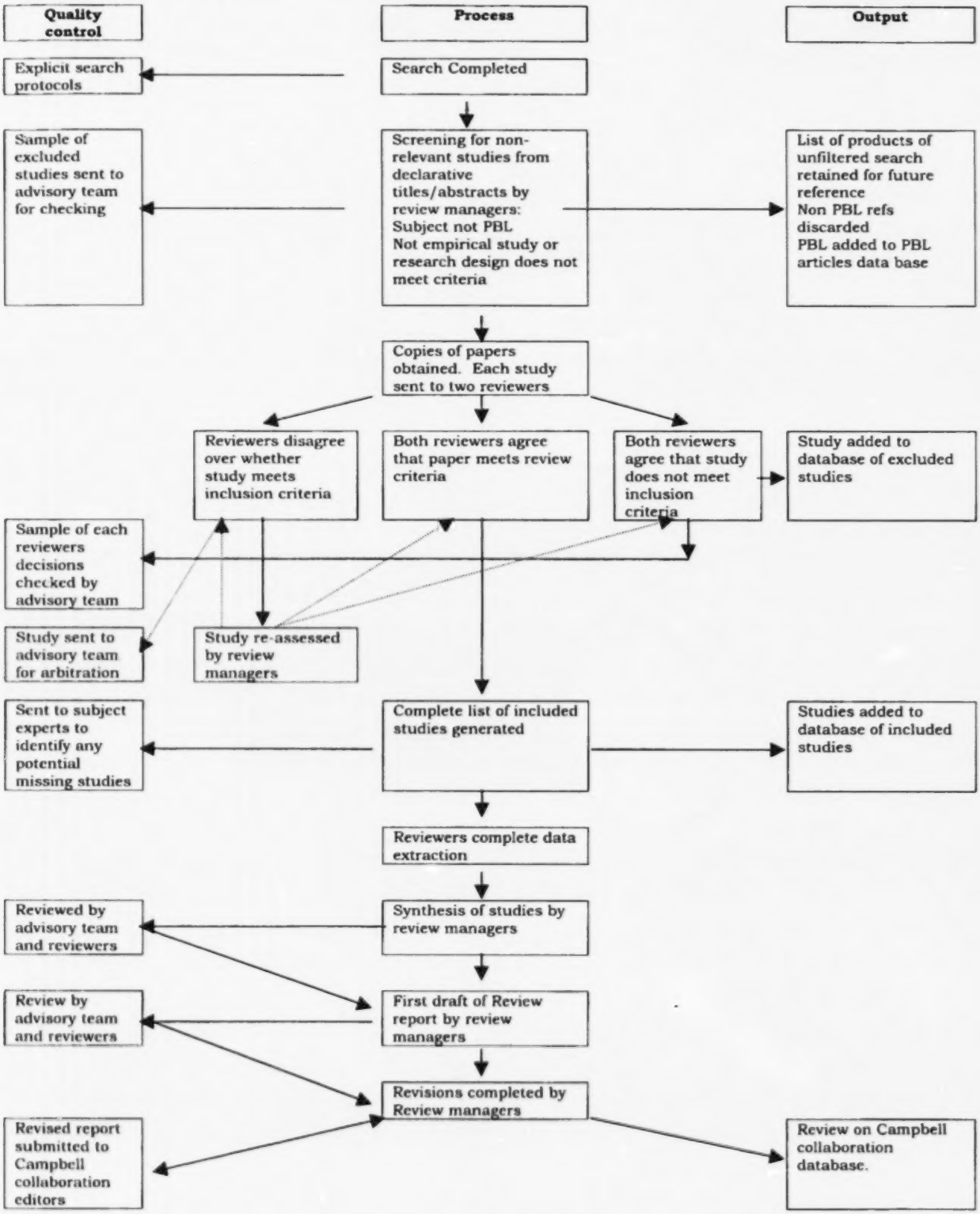
The review process is outlined in figure 1. The products generated by the search strategy will take the form of full reference and abstract. These will be entered onto a database and passed to the review managers. The review managers will independently screen these lists to identify papers that appear to meet the methodological inclusion criteria of the review i.e. papers in which the authors state that the research design is one of the types required by the review protocol. Full text copies of all papers which either of the review managers consider should be included will be obtained.

Depending on the number of full text articles generated at this stage two alternative pathways may be followed.

i) If a large number of studies are identified for potential inclusion the review managers will screen the full text papers to identify those interventions which appear to be 'authentic PBL' using the review criteria. Copies of the full text of articles where the educational intervention is 'authentic PBL' will be distributed amongst the study quality assessment panel. Articles where the educational intervention is hybrid or combination PBL curricula will be placed on file for future rounds of the review.

ii) If the number of potential studies is sufficiently small then copies of all the full text articles will be distributed amongst the study quality assessment panel

Figure 1: Review process


Quality control
Process
Output
Explicit search protocols
Search Completed
Sample of excluded studies sent to advisory team for checking
Screening for non-relevant studies from declarative titles/abstracts by review managers:
Subject not PBL
Not empirical study or research design does not meet criteria
List of products of unfiltered search retained for future reference
Non PBL refs discarded
PBL added to PBL articles data base
Copies of papers obtained. Each study sent to two reviewers
Reviewers disagree over whether study meets inclusion criteria
Both reviewers agree that paper meets review criteria
Both reviewers agree that study does not meet inclusion criteria
Study added to database of excluded studies
Sample of each reviewers decisions checked by advisory team
Study sent to advisory team for arbitration
Study re-assessed by review managers
Sent to subject experts to identify any potential missing studies
Complete list of included studies generated
Studies added to database of included studies
Reviewers complete data extraction
Reviewed by advisory team and reviewers
Synthesis of studies by review managers
Review by advisory team and reviewers
First draft of Review report by review managers
Revisions completed by Review managers
Revised report submitted to Campbell collaboration editors
Review on Campbell collaboration database.

3.2 Study quality assessment panel

Membership of the review quality assessment panel is given in appendix 4. All members of the panel have experience of or training in systematic review methods. The panel members have a different balance of experience between PBL and systematic reviewing. Each study will be sent to two reviewers for quality assessment and data extraction. A panel member who is more experienced in PBL and a panel member who is more experienced in systematic reviewing will review each study. Panel members will not review studies for which they are a reported author. Quality assurance mechanisms and procedures for resolving disagreement between panel members are outlined in figure 1.

3.3 Inclusion criteria

3.3.1 Population:

The review will only include participants in post-school education programmes.

Score DONE if study participants are undertaking educational programmes in: tertiary, college, university, Further or adult education. Undergraduate or post-graduate programmes. Professional, vocational, post-registration or post-qualification training. Personal or professional development programmes.

Score NOT CLEAR if participant age not given(N.B. the paper should be discussed with the review managers before data extraction is undertaken).

Score NOT DONE if age range (setting) of participants not given and clearly not obtainable or participants are school age children.

The following characteristics of participants/setting will be reported: Country, profession, subject, academic level of course/training, professional specialty, age, time since graduation, setting, incentives/motivation, academic status of programme.

If you scored NOT DONE for the above criteria the study should not be included in the review.

3.3.2 Study designs

The review will use the standard EPOC criteria for study designs. These are Randomized Controlled Trials (RCT), Controlled Clinical Trials (CCT) , Interrupted Time Series (ITS), Controlled Before & After studies (CBA). Qualitative data collected within such studies e.g. researchers observations of events, will be incorporated in reporting. Studies that utilize solely qualitative approaches will not be included in the review.

The design of the study is (state which):

Randomised controlled trial (RCT) i.e. a trial in which the participants (or other units) were definitely assigned prospectively to one or two (or more) alternative forms of health care using a process of random allocation (e.g. random number generation, coin flips).

Controlled clinical trial (CCT) may be a trial in which participants (or other units) were:

a) definitely assigned prospectively to one or two (or more) alternative forms of health care using a quasi-random allocation method (e.g. alternation, date of birth, student identifier) or;

b) possibly assigned prospectively to one or two (or more) alternative forms of education using a process of random or quasi-random allocation.

Controlled before and after study (CBA) i.e. involvement of intervention and control groups other than by random process, and inclusion of baseline period of assessment of main outcomes. There are two minimum criteria for inclusion of CBAs in the review:

a) Contemporaneous data collection
Score DONE pre and post intervention periods for study and control sites are the same.
Score NOT CLEAR if it is not clear in the paper, e.g. dates of collection are not mentioned in the text. (N.B. the paper should be discussed with review managers before data extraction is undertaken).
Score NOT DONE if data collection was not conducted contemporaneously during pre and post intervention periods for study and control sites.

b) Appropriate choice of control site:
Studies using second site as controls:
Score DONE if study and control sites are comparable with respect to dominant reimbursement system, subject/discipline/professional group, academic status/level of programme.
Score NOT CLEAR if not clear from paper whether study and control sites are comparable. (N.B. the paper should be discussed with the review managers before data extraction is undertaken).
Score NOT DONE if study and control sites are not comparable.

Interrupted time series (ITS) i.e. a change in trend attributable to the intervention. There are two minimum criteria for inclusion of ITS designs in the review:

a) Clearly defined point in time when the intervention occurred.
Score DONE if reported that intervention occurred at a clearly defined point in time.
Score NOT CLEAR if not reported in the paper (will be treated as NOT DONE if information cannot be obtained from the authors).
Score NOT DONE if reported that intervention did not occur at a clearly defined point in time.

b) At least three data points before and three after the intervention.
Score DONE if 3 or more data points before and 3 or more data points recorded after the intervention. Score NOT CLEAR if not specified in paper e.g. number of discrete data points not mentioned in text or tables (will be treated as NOT DONE if information cannot be obtained from the authors).
Score NOT DONE if less than 3 data points recorded before and 3 data points recorded after intervention.

If the study is not any of the above designs, it should not be included in the review. If you scored NOT DONE for any of the above criteria in 3.1.2, the study should not be included in the review. If reviewers are unsure of the study design, the paper should be discussed with the review managers before data extraction is undertaken.

3.3.3 Methodological inclusion criteria

The minimum methodological inclusion criteria across all study designs are:

a) The objective measurement of student performance/behaviour or other outcome(s).
Score DONE (e.g. performance of providers against pre-set criteria, or in performance assessment). Outcome measures such as student satisfaction with work or student satisfaction with care may be included if they are assessed using a standard systematic technique/ approach.
Score NOT CLEAR (the paper should be discussed with the review managers before data extraction is undertaken).
Score NOT DONE (e.g. self-reported data, measurement of attitudes, beliefs, perceptions or satisfaction).

b) Relevant and interpretable data presented or obtainable.
Score DONE if data was presented or obtainable.
Score NOT CLEAR (the paper should be discussed with the review managers before data extraction is undertaken).
Score NOT DONE if relevant data was not presented and is clearly unobtainable.

If either of the above criteria is scored as NOT DONE, the study should not be included in the review.

3.3.4 Type of Intervention:

The minimum inclusion criteria for interventions for the initial review are:

A) Cumulative integrated curriculum (in the case of single subject educational programmes that do not have a explicit professional aspect e.g. English Literature, or short post- registration programmes integration may not be a requirement)

Score DONE if sufficient description of curriculum given and integration present and/or not required/relevant.

Score NOT CLEAR if inadequate description or where PBL is used on one module within otherwise non-pbl programme (the paper should be discussed with the review managers before data extraction is undertaken)

Score NOT DONE if not sufficient description of curriculum and is clearly not obtainable or if curriculum is not integrated or if PBL is used in single module/ pathway of curriculum that is not separately assessed (and for which data is presented in the particular study under review)

B) Learning via simulation formats that allow free enquiry (i.e. not problem solving learning)

Score DONE if Problem/trigger/scenario presented prior to and for the purpose of learning

Score NOT CLEAR if inadequate description or where lectures are also used within the PBL programme (The paper should be discussed with the review managers before data extraction is undertaken)

Score NOT DONE if there is insufficient description of the simulation format and/or it is clearly not obtainable and/or Problem/trigger/scenario is presented to practice learning undertaken through other formats

C) Small groups with either faculty or peer tutoring

Score DONE if learning is undertaken in small groups (<20 students)

Score NOT CLEAR if inadequate description (the paper should be discussed with the review managers before data extraction is undertaken)

Score NOT DONE if there is insufficient description and/or it is clearly not obtainable or where the number of students in the same teaching and learning session is > 20

D) An explicit framework is followed in tutorials e.g. Maastricht 7 steps, that includes as a minimum: Exploration of the potential issues/hypothesis, identfication of learning needs, student activity to meet the identified learning needs, group feedback and discussion of the information obtained, elaboration through the application of the new knowledge to the scenario, review of learning and group process.

Score DONE if clear description of procedural steps in tutorial process are given or reference made to use of established procedure e.g. Barrows tutorial process.

Score NOT CLEAR if inadequate description or where not clear whether the procedures in tutorial process include all the appropriate stages (the paper should be discussed with the review managers before data extraction is undertaken)

Score NOT DONE if there is insufficient description and/or it is clearly not obtainable or if clear that a tutorial process not used and/or does not include all the relevant stages.

If any of the above criteria is scored as NOT DONE, the study should not be included in the review. If reviewers are unsure of the intervention design, the paper should be discussed with the review managers before data extraction is undertaken.

The criteria given above will be used to identify studies for inclusion in the review. For included studies a full description of the PBL intervention will be given using Barrows descriptive criteria (see appendix 1). Only studies that meet all of the criteria above will be included in the analysis for the initial review. Studies that meet

the research methodology criteria and more than one but not all of the above criteria (i.e. maybe considered a hybrid or combination PBL curricula) will be included in the database for analysis of the secondary review question.

4 Control groups.

Control groups will be categorized as follows

a) No intervention control group
b) Standard practice control group (if different to (a))
c) Untargeted activity
d) Other (e.g. other intervention)

5 Quality Assessment of primary studies

5.1. Quality of study design

The quality assessment criteria are designed to exclude studies where there is a significant risk of bias due to weakness in the study design. Each criterion is scored as done, Not Clear, Not done. Detailed descriptions of the criteria are given in the data collection checklist (see appendix 2)

The criteria for RCTs and CCTs are:

- Concealment of allocation
- Follow-up of participants
- Blinded assessment of primary outcomes (scored as done if outcomes were assessed blindly or the outcome variables are objective)
- Baseline measurement
- Protection against contamination

The Criteria for CBA studies are:

- Baseline measurement
- Baseline characteristics
- Blinded assessment of primary outcome measures (see above)
- Protection against contamination
- Reliable primary outcome measure
- Follow up of professionals

The criteria for ITS studies are:

- Intervention independent of other changes
- Sufficient data points to enable reliable statistical inference (number of points required for different methods of analysis given)
- Formal test for trend
- Intervention unlikely to affect data collection
- Blinded assessment of primary outcomes (see above)
- Completeness of data se
- Reliable primary outcome measures

5.2 Power calculation

Score DONE if study has sufficient statistical power to detect educationally important effects as statistically significant and record power.
Score NOT CLEAR if not reported.
Score NOT DONE if authors specifically report that the study was under-powered

6 OUTCOMES

6.1 Description of the main outcome measure(s)

A variety of outcome measures may be reported in studies of PBL. Friedman et al (1990) highlight a number of hypothesized areas of difference between traditional and problem-based and/or community-oriented curricula in the education of health professionals. Those that could be interpreted as outcomes produced by PBL are reproduced in box 1. They hypothesize that traditional curricula will produce better results in board examinations and other tests of knowledge which may be problematic as many PBL studies are likely to use these kinds of outcome measures.

The review will however report all the main outcomes described by the authors.

a) Student/faculty outcomes/process measures (e.g. attainment scores, satisfaction)

b) patient/client outcomes (e.g. the number of adverse drug events)

c) Economic variables

- Costs of the intervention:
  Score DONE if reported, and describe costs;
  Score NOT DONE if not reported

- Changes in practice as a result of the intervention (e.g. improved assessment scores etc):
  Score DONE if reported, and describe costs;
  Score NOT DONE if not reported

- Costs associated with the intervention are linked with student or service user outcomes in an economic evaluation (e.g. net cost per unit of change in assessment performance):
  Score DONE if reported, and describe ratio;
  Score NOT CLEAR if not adequately described in the paper;
  Score NOT DONE if there was no economic evaluation reported.

6.2 Length of time during which outcomes were measured after initiation of the intervention

6.3 Length of post-intervention follow-up period

Score DONE if reported in the paper (specify length of follow-up period)
Score NOT CLEAR if not reported in the paper
Score NOT DONE if there was no follow-up period.

6.4 Identify a possible ceiling effect

For example, there was little room for improvement in student performance, because it was adequate without the intervention (based on baseline measurements or control group performance).

a) Identified by investigator
- Yes
- No
- NOT CLEAR

b) Identified by reviewer
- Yes
- No
- NOT CLEAR

7. RESULTS

State the results as they will be entered in the review, and describe how these were calculated (e.g. relative percentage differences attributable to the intervention).

7.1 For RCTs and CCTs

a) State the main results of the main outcome(s), for each study group, in natural units.

b) For each available comparison, report the baseline and post intervention differences between study and control groups, in natural units. Include statistical significance if reported. Indicate if the unit of randomisation and analysis were different.

In all cases, report a more favourable provider/student outcome in the more active intervention group as a positive (+) finding (i.e. where differences in the groups are in the intended direction).

7.2 For CBAs

a) State the main results of the main outcome(s), for each study group, in natural units.

b) For each study group, report baseline and post intervention results. Calculate the pre-post intervention difference for each outcome in natural units (i.e. the post-intervention outcome minus the pre-intervention outcome).

c) For each available comparison, calculate the difference across study groups of the pre-post intervention change (i.e. if, for an outcome measure ΔE is the pre-post intervention change in the experimental/intervention group, and ΔC is the pre-post intervention change in the control group, this will be $\Delta E - \Delta C$).

Include statistical significance if reported.

In all cases, report a more favourable provider/student outcome in the more active intervention group as a positive (+) finding (i.e., where differences in the groups are in the intended direction).

7.3 For ITSs

State the main results of the main outcome(s) in natural units.

In all cases, report a more favourable provider/student outcome attributable to the intervention as a positive (+) finding (i.e. where changes in the outcomes are in the intended direction).

8 Broad strategy for searching

The searching will be carried out by an information scientist familiar with the principles of systematic reviewing and searching of bibliographic databases for this purpose. The resources available will limit the extent of the search. All stages of the search process will be documented and highlighted and identified deficiencies reported. It is acknowledged that subjects and disciplines outside medicine and the health sciences may be under represented in the review as the team lack detailed knowledge of sources in these areas. It is likely that a number of relevant papers will not be identified in the database search (Petrosino 2000, Personal communication Alex Haig, information scientist BEME). However hand searching of journals will not be possible in the initial review. It is hoped that hand searching may take place in subsequent review updates.

8.1 Databases to be searched

MEDLINE, CINAHL, ERIC, Research & Development Resource Base in Continuing Medical Education, EMBASE, HealthSTAR, EPOC specialist register, Cochrane Controlled Trials Register (CCTR), Psychinfo, TIMELIT (Topics In Medical Education), C2-SPECTR (Social, Psychological, Educational and Criminological Trials), Higher Education Research Abstracts, British Education Index . Psychological abstracts, Sociological Abstracts.

8.2 Key words

The initial search of each of the databases indicated above will use the list of key words below, generated in consultation with PBL experts worldwide

Problem-based learning, Problem based learning, PBL (with and without capitalization), Enquiry -based learning, Enquiry based learning, EBL, Inquiry-based learning, Inquiry based learning, IBL, Problem centered learning, small group learning, problem solving learning, active learning, co-operative learning, practice based learning, self directed learning, learner centered learning, action learning, community based learning, scenario based learning, work based learning, student centered learning, guided learning, emancipatory learning, capability learning, case-based learning, Inquiry approaches, anchored instruction, cognitive apprenticeship, issues based learning, task-based learning, problem first learning, self-discovery learning, self-managed learning, learner managed learning, peer group learning, ASSET model, learning sets, case led learning, contextual learning, resource based learning, situated learning, shared learning, experiential learning, learning through work, enquiry and action learning, group learning, collaborative learning/enquiry, negotiated learning, transformatory learning, case centred learning.

Plus : substitute word curriculum for learning in above examples
Plus: Use term 'hybrid' as often used as description of combined PBL course
Plus: substitute tem 'development' for learning

8.3 Search Inclusion/Exclusion criteria:

Post- school, Education method, no limitation by discipline, date. Data will be extracted from any papers that meet review criteria whatever the subject discipline of the students.
Decisions about synthesis will be taken in the light of the search results. For MEDLINE, HealthSTAR and EMBASE study methodology filters (so called expert filters) will be applied to the set of references identified using the key word search to identify studies meeting the review entry criteria. Modified versions of the filters developed by the Cochrane group on Effective Practice and Organization of Care (EPOC) will be used (see appendix 3)

8.4 Assessing sensitivity of search strategies

Products of the search will be entered onto a database and be made available electronically. Studies identified in existing published reviews of PBL 'reviews' will be used as a 'gold standard' against which the sensitivity of the search strategy will be assessed both before and after the application of methodological filters.

8.5 Citation searching:

Science Citation index, Social Science Citation index, review of reference lists

8.6 Identification of grey literature:

i) The main strategy for the identification of grey literature is the contact of PBL experts and checking of reference lists. A basic search of the worldwide web search will also be completed.

9 Data Extraction

A standard data extraction sheet will be used by all reviewers to extract information and data from the individual study reports. Where data is missing from study reports attempts will be made to contact the study authors.

10 Data synthesis

If possible outcome data will be synthesized for all included studies. Depending on the number and characteristics of the studies identified subgroup analysis by subject or discipline may also be undertaken. There are two principle approaches to data synthesis. Non quantitative synthesis will involve tabulation of study characteristics and results to summarize their findings. This approach allows a qualitative assessment of the evidence. The principle objective of this analysis is to summarize the information about the included studies in a relevant and meaningful way. The synthesis will highlight similarities and differences between studies also whether studies are similar enough to calculate average estimates of effectiveness. Where the it is not possible to carry out quantitative analysis it may still be possible to estimate a qualitative effect through inclusion of authors comments and/or the use of techniques such as vote counting.

This will be supplemented, where possible, by quantitative synthesis using statistical methods to assess variation in results and generate pooled results if appropriate. Quantitative analysis is only possible if the required numerical data is available and there is sufficient homogeneity between studies. It is likely that studies of PBL will use as outcome measures different types of data. The studies in the Vernon and Blake (1993) review for example include outcomes such as assessment scores, quality ratings and teaching method preferences. These outcomes may be treated as continuous, dichotomous or ordinal data. Analysis will focus on comparing differences in effect between the intervention and control.

Depending on the measurement scale of the outcomes an effect measure will be generated as a change in event rate or as a change on a continuous scale. For event data changes will be measured in terms of relative and absolute differences. For continuous data the effect data are based on differences in means or standardized mean differences (Effect sizes). The data from individual studies and any meta-analysis will be presented using forest plots where odds ratios or relative risk are used an effect measure. Where effect sizes are used presentation of differences in the distribution curves will be explored.

Heterogeneity between studies will be assessed using visual methods (L'Abbe plots) and formal statistical tests. The synthesis of different study designs will be conducted either by analysing each study design as a separate group or by cumulatively combining studies with decreasing strengths of evidence. Sensitivity analysis will be used to test the robustness of the results of the meta-analysis. Formal estimates of the effect of publication bias will be carried out using a funnel plot to show the distribution of effect sizes according to sample size. Large gaps in the funnel indicate the possibility of missing publications.

11 Timetable

From 2002	Feb	Mar	apr	may	june	july	Aug	Sept	oct	nov	dec	Jan
Medline search	→	→										
Other search strategies			→	→								
Assess studies from reviews citation list		→	→									
Develop review forms		→	→									
Screening from reference lists			→	→	→							
Review and data extraction					→	→	→	→				
Synthesis & Analysis							→	→	→			
Draft report to reviewers										■		
Revisions											■	
Publication of final report												■

12 Dissemination strategy

M. Newman is currently the principle investigator on project funded by the ESRC Teaching & Learning Research Programme on the effectiveness of PBL. As part of this project an international database of PBL contacts, a PBL e-mail discussion group, a PBL website, and a PBL newsletter have already been established. Two user engagement events are also planned in London in 2001 and 2002. Links have been established with other relevant research networks and the HEFCE funded Learning & Teaching support network in the UK. Dissemination will take place through all of these networks as well as through the Campbell collaboration and publication in peer and professional journals.

Criteria For analyzing a problem based learning curriculum (Barrows 2000b)

1) Over what time span does the PBL curriculum occur
2) Is the PBL experience designed for all students or for a subset (alternative or parallel curriculum)
3) How is the time for PBL distributed over the calendar e.g. hours day , 1 day week
4) Is the PBL an add-on to the regular curriculum without modification or reduction of what students have always had to learn in the regular curriculum
5) How many subjects are integrated into the PBL experience for the students
6) What subject are taught outside of PBL and what methods are used to teach those subjects
7) If PBL occurs throughout a block, course, year, are there other teaching/learning methods used outside of the PBL experience (lectures, laboratories), if there are other teaching methods are they integrated into the PBL curriculum, Are they optional for the student to take/ attend?
8) Is the PBL experience designed for small groups of students or the whole class
9) If it is designed for small groups, how many students are in the group
10) Are one or two tutors used? If more than one is used what are their respective roles
11) What are the background of the tutors (Are they subject experts in the substantial part of the programme for which they act as PBL tutors?)
12) What kind of training is used to prepare the tutors
13) How are the tutors evaluated and by whom
14) Which of the following stages in the PBL process are used by tutors

- reasoning through the problem using the hypothetico – deductive process
- Identification of learning issues
- Identification of information resources to be employed during self-directed learning
- Critique of learning resources on return from self – study
- Application of newly acquired information back to the problem.
- Discussion of learning to achieve integration, abstraction and transfer
- Self assessment by students
- Peer assessment by students

15) How are the problems designed (Is a statement of task or question given in the scenario or trigger material)
16) Do the problem stimulation formats used permit students to freely inquire about findings on history, physical and laboratory tests
17) Are there experiences with standardized and /or real Patients/clients incorporated into the PBL experience
18) What is the role of resource faculty or other teaching faculty outside the tutor?
19) How are students assessed
20) Are they given grades or pass/fail scores and how are these determined.

Quality criteria for randomised controlled trials (RCTs & CCTs)

Seven standard criteria are used for randomised controlled trials and controlled clinical trials included in the review:

a) Concealment of allocation (protection against selection bias)
Score DONE if
- the unit of allocation was by institution, group, or teacher and any random process is described explicitly, e.g. the use of random number tables or coin flips;
- the unit of allocation was by student and there was some form of centralised randomisation scheme, an on-site computer system or sealed opaque envelopes were used.

Score NOT CLEAR if
- the unit of allocation is not described explicitly;
- the unit of allocation was by student and the authors report using a 'list' or 'table', 'envelopes' or 'sealed envelopes' for allocation.

Score NOT DONE if
- the authors report using alternation such as reference to case record numbers, dates of birth, day of the week or any other such approach (as in CCTs);
- the unit of allocation was by student and the authors report using any allocation process that is entirely transparent before assignment such as an open list of random numbers or assignments;
- allocation was altered (by investigators, teachers or students).

b) Follow-up of professionals (protection against exclusion bias)
Score DONE if outcome measures obtained for 80-100% of subjects randomised. (Do not assume 100% follow up unless stated explicitly.);
Score NOT CLEAR if not specified in the paper;
Score NOT DONE if outcome measures obtained for less than 80% of subjects randomised.

NB: Course drop-out rate may itself be used as a study outcome measure

c) Follow-up of students
Score DONE if outcome measures obtained for 80-100% of subjects randomised or for students who entered the trial. (Do not assume 100% follow up unless stated explicitly.) Score DONE if there is an objective data collection system;
Score NOT CLEAR if not specified in the paper;
Score NOT DONE if outcome measures obtained for less than 80% of subjects randomised or for less than 80% of students who entered the trial.

d) Blinded assessment of primary outcome(s)* (protection against detection bias)
Score DONE if the authors state explicitly that the primary outcome variables were assessed blindly OR the outcome variables are objective, e.g. assessment using a standardised test;
Score NOT CLEAR if not specified in the paper;
Score NOT DONE if the outcome(s) were not assessed blindly.

* ***Primary outcome(s) are those variables that correspond to the primary hypothesis or question as defined by the authors. In the event that some of the primary outcome variables were assessed in a blind fashion and others were not, score each separately and label each outcome variable clearly.***

e) Baseline measurement
Score DONE if performance or student outcomes were measured prior to the intervention, and no substantial differences were present across study groups;

Score NOT CLEAR if baseline measures are not reported, or if it is unclear whether baseline measures are substantially different across study groups;
Score NOT DONE if there are differences at baseline in main outcome measures likely to undermine the post intervention differences (e.g. are differences between the groups before the intervention similar to those found post intervention).

f) Reliable primary outcome measure(s)*
Score DONE if two or more raters with at least 90% agreement or kappa greater than or equal to 0.8 OR the outcome is obtained from some automated system e.g. course completion rates, assessment using a standardised test;
Score NOT CLEAR if reliability is not reported for outcome measures that are obtained by assignments or collected by an individual;
Score NOT DONE if agreement is less than 90% or kappa is less than 0.8.

* ***In the event that some outcome variables were assessed in a reliable fashion and others were not, score each separately on the back of the form and label each outcome variable clearly.***

g) Protection against contamination
Score DONE if allocation was by community, institution or practice and it is unlikely that the control received the intervention;
Score NOT CLEAR if professionals were allocated within the same institution and it is possible that communication between the experimental and control group could have occurred;
Score NOT DONE if it is likely that the control group received the intervention (e.g. cross-over trials or if the same teachers taught both control and intervention groups).

Quality criteria for controlled before and after (CBA) designs

Seven standard criteria are used for CBAs included in the review:

a) Baseline measurement
Score DONE if performance or student outcomes were measured prior to the intervention, and no substantial differences were present across study groups (e.g. where multiple pre intervention measures describe similar trends in intervention and control groups);
Score NOT CLEAR if baseline measures are not reported, or if it is unclear whether baseline measures are substantially different across study groups;
Score NOT DONE if there are differences at baseline in main outcome measures likely to undermine the post intervention differences (e.g. are differences between the groups before the intervention similar to those found post intervention).

b) Characteristics for studies using second site as control
Score DONE if characteristics of study and control providers are reported and similar;
Score NOT CLEAR if it is not clear in the paper e.g. characteristics are mentioned in the text but no data are presented;
Score NOT DONE if there is no report of characteristics either in the text or a table OR if baseline characteristics are reported and there are differences between study and control providers.

c) Blinded assessment of primary outcome(s)* (protection against detection bias)
Score DONE if the authors state explicitly that the primary outcome variables were assessed blindly OR the outcome variables are objective e.g. course completion rates or assessment by a standardised test;
Score NOT CLEAR if not specified in the paper;
Score NOT DONE if the outcomes were not assessed blindly.

* ***Primary outcome(s) are those variables that correspond to the primary hypothesis or question as defined by the authors. In the event that some of the primary outcome variables were assessed in a blind fashion and others were not, score each separately and label each outcome variable clearly.***

d) Protection against contamination
Studies using second site as control
Score DONE if allocation was by community, institution, or practice and is unlikely that the control group received the intervention;
Score NOT CLEAR if providers were allocated within the same institution or practice and communication between experimental and group providers was likely to occur;
Score NOT DONE if it is likely that the control group received the intervention (e.g. cross-over studies or if the same teachers taught both control and intervention groups).

e) Reliable primary outcome measure(s)
Score DONE if two or more raters with at least 90% agreement or kappa greater than or equal to 0.8 OR the outcome is obtained from some automated system e.g. course completion rates, assessment using a standardised test;
Score NOT CLEAR if reliability is not reported for outcome measures that are based on assignment marks or are collected by an individual;
Score NOT DONE if agreement is less than 90% or kappa is less than 0.8.

* ***In the event that some outcome variables were assessed in a reliable fashion and others were not, score each separately and label each outcome variable clearly.***

f) Follow-up of students (protection against exclusion bias)
Score DONE if outcome measures obtained 80-100% subjects allocated to groups. (Do not assume 100% follow-up unless stated explicitly.);
Score NOT CLEAR if not specified in the paper;
Score NOT DONE if outcome measures obtained for less than 80% of students allocated to groups.

g) Follow-up of students
Score DONE if outcome measures obtained 80-100% of students allocated to groups or for students who entered the study. (Do not assume 100% follow-up unless stated explicitly.);
Score NOT CLEAR if not specified in the paper;
Score NOT DONE if outcome measures obtained for less than 80% of students allocated to groups or for less than 80% of students who entered the study.

6.4.3 *Quality criteria for interrupted time series (ITSs)*

The following seven standard criteria should be used to assess the methodology quality of ITS designs included in the reviews. Each criterion is scored DONE, NOT CLEAR or NOT DONE.

a) Protection against secular changes

- The intervention is independent of other changes.
  Score DONE if the intervention occurred independently of other changes over time;
  Score NOT CLEAR if not specified (will be treated as NOT DONE if information cannot be obtained from the authors);
  Score NOT DONE if reported that intervention was not independent of other changes in time.
- There are sufficient data points to enable reliable statistical inference
  Score DONE
  (a) if at least twenty points are recorded before the intervention **AND** the authors have done a traditional time series analysis (ARIMA model) (or a post hoc analysis can be done)
  OR

repeated measures analysis (or a post hoc analysis can be done)

OR

(c) if at least 3 points are recorded pre and post intervention **AND** the authors have used ANOVA or multiple t-tests (or a post hoc analysis can be done) AND there are at least 30 observations per data point.

Score NOT CLEAR if not specified in paper e.g. number of discrete data points not mentioned in text or tables (will be treated as NOT DONE if information cannot be obtained from the authors);

Score NOT DONE if any of the conditions above are unmet.

- Formal test for trend. Complete this section if authors have used ANOVA modelling.
  Score DONE if formal test for change in trend using appropriate method is reported (e.g. see Cook & Campbell 1979[1]) (or can be re-done);
  Score NOT CLEAR if not specified in paper (will be treated as NOT DONE if information cannot be obtained from the authors);
  Score NOT DONE if formal test for change in trend has not been done.

b) Protection against detection bias

- Intervention unlikely to affect data collection
  Score DONE if reported that intervention itself was unlikely to affect data collection (for example, sources and methods of data collection were the same before and after the intervention);
  Score NOT CLEAR if not reported (will be treated as NOT DONE if information cannot be obtained from the authors);
  Score NOT DONE if the intervention itself was likely to affect data collection (for example, any change in source or method of data collection reported).
- Blinded assessment of primary outcome(s)*
  Score DONE if the authors state explicitly that the primary outcome variables were assessed blindly OR the outcome variables are objective e.g. drop out rates or assessment by a standardised test;
  Score NOT CLEAR if not specified (will be treated as NOT DONE if information cannot be obtained from the authors);
  Score NOT DONE if the outcomes were not assessed blindly.

* ***Primary outcome(s) are those variables that correspond to the primary hypothesis or question as defined by the authors. In the event that some of the primary outcome variables were assessed in a blind fashion and others were not, score each separately and label each outcome variable clearly.***

c) Completeness of data set
Score DONE if data set covers 80-100% of total number of participants or episodes of care in the study;
Score NOT CLEAR if not specified (will be treated as NOT DONE if information cannot be obtained from the authors);
Score NOT DONE if data set covers less than 80% of the total number of participants in the study.

d) Reliable primary outcome measure(s)*
Score DONE if two or more raters with at least 90% agreement or kappa greater than or equal to 0.8 OR the outcome is obtained from some automated system e.g. course completion rates, assessment using a standardised test;
Score NOT CLEAR if reliability is not reported for outcome measures that are based on assignment marks or are collected by an individual (will be treated as NOT DONE if information cannot be obtained from the authors);
Score NOT DONE if agreement is less than 90% or kappa is less than 0.8.

#revised EPOC strategy Medline & Health star- 31/01/00

randomized controlled trial.pt.
controlled clinical trial.pt.
intervention studies/
experiment$.tw.
(time adj series).tw.
(pre test or pretest or (posttest or post test)).tw.
random allocation/
impact.tw.
intervention?.tw.
chang$.tw.
evaluation studies/
evaluat$.tw.
effect?.tw.
comparative studies/
animal/
human/
119 not 120
or/105-118
122 not 121
104 and 123

EPOC EMBASE search strategy

1. randomized controlled trial/
2. (randomised or randomized).tw.
3. experiment$.tw.
4. (time adj series).tw.
5. (pre test or pretest or post test or posttest).tw.
6. impact.tw.
7. intervention?.tw.
8. chang$.tw.
9. evaluat$.tw.
10. effect?.tw.
11. compar$.tw.
12. or/1-11
13. nonhuman/
14. 12 not 13

Full reference for paper for being reviewed (please write in)

Section One: Inclusion criteria

Criterion	Done	Not Done	Not Clear	
Study design: RCT Assigned prospectively using a process of random allocation				Go to Section 2.1
Study design: CCT Definitely or possibly assigned prospectively using a quasi-random allocation method				Go to section 2.1
Study design: CBA a) Contemporaneous data collection b)Appropriate choice of control site:				Go to section 2.2
Study design: ITS a) Clearly defined point in time when the intervention occurred. b) At least three data points before and three after the intervention.				Go to section 2.3
Methodological inclusion criteria a) Objective measurement of student performance/ behaviour or other outcome(s). b) Relevant and interpretable data presented or obtainable.				Go to Sections 2 & 4
Population: Post compulsory school age programmes				Go to section 3
Type of Intervention: a) Cumulative integrated curriculum (where appropriate) b) Learning via simulation formats that allow free enquiry c) Small groups with either faculty or peer tutoring d) Explicit framework followed in tutorials				Go to Section 5

If you scored NOT DONE for any of the above criteria the study should not be included in the review. If reviewers are unsure or have marked any criterion 'unclear the paper should be discussed with the review managers before data extraction is undertaken.

Section 2 Quality Assessment of primary studies (paper ref. first author & year_____)

2.1 Quality criteria for randomised controlled trials (RCTs & CCTs)

Criterion	Done	Not Done	Not clear	Notes	Criterion	Done	Not Done	Not clear	Notes
Concealment of allocation (protection against selection bias)					Baseline measurement				
Follow-up of students (protection against exclusion bias)					Protection against contamination				
Power calculation									

	Blinded assessment			**Reliability**			**Results** – Main outcome for each group in natural units			**Results** pre-post change in intervention and control groups		
Primary outcome(s) (Score each outcome separately)	Done	Not done	Not Clear	Done	Not done	Not Clear	Intervention group (1)	Control group (2)	(1)-(2)	Intervention group change (3)	Control group change (4)	(3) –(4)

Possible ceiling effect	Yes	No	Not clear	**Indicate which if any results statistically significant and /or where unit of analysis is different from unit of randomization:**
Identified by investigator				
Identified by reviewer				

Section 2 Quality Assessment of primary studies (paper ref. first author & year_____)

2.2 Quality criteria for controlled before and after designs (CBA)

Criterion	Done	Not Done	Not clear	Notes	Criterion	Done	Not Done	Not clear	Notes
Characteristics of control site					Baseline measurement				
Follow-up of students (protection against exclusion bias)					Protection against contamination				
Power calculation									

	Blinded assessment			Reliability			Results – Main outcome for each group in natural units			Results pre-post change in intervention and control groups		
Primary outcome(s) (Score each outcome separately)	Done	Not done	Not Clear	Done	Not done	Not Clear	Intervention group (1)	Control group (2)	(1)-(2)	Intervention group change (3)	Control group change (4)	(3) –(4)

Possible ceiling effect	Yes	No	Not clear	**Indicate which if any results statistically significant**
Identified by investigator				
Identified by reviewer				

Section 2 Quality Assessment of primary studies (paper ref. first author & year_____)

2.3 Quality criteria for interrupted time series designs (ITS)

Criterion	Done	Not Done	Not clear	Notes	Criterion	Done	Not Done	Not clear	Notes
Protection against contamination					Power calculation				
Intervention unlikely to affect data collection					Baseline measurement				
Protection against secular changes					Completeness of dataset				
Intervention independent of other changes									
Sufficient data points									
Formal test for trend									

	Blinded assessment			Reliability			Results – Main outcome for each group in natural units
Primary outcome(s) (Score each outcome separately)	Done	Not done	Not Clear	Done	Not done	Not Clear	

Possible ceiling effect	Yes	No	Not clear
Identified by investigator			
Identified by reviewer			

3. Population: (paper ref. first author & year_____)

3.1 Control groups.

(1) No intervention control group	(2) Standard practice control group (if different to (1)	(3)Untargeted activity	(4) Other (e.g. other intervention)

3.2 Sample Characteristics

Characteristic	**Intervention group**	**Control group**
Country of study		
Method of entry to educational programme (e.g. criteria, selection)		
Profession/ discipline (e.g. medicine/ sociology		
Subject of study e.g. neurophysiology/ post-modernism		
Academic level of course (e.g. B.A, MA) (state year where appropriate)		
Age (average – min & max)		
Time since graduation (average – min & max)		
Setting (e.g. Adult education college)		
Incentive/ motivation (e.g. programme tied to promotion or payrise)		

4 Outcome measures

	Done	**Not Done**	**Not clear**	**Description**
Economic variables • Costs of the intervention • Changes in practice as result of intervention • Costs linked with student or service user outcomes				
Length of post intervention follow-up period				
Length of time during which outcomes were measured from start of intervention period				

5 Describing the intervention and control

Descriptors	Intervention (PBL curriculum)	Control
Over what time span does the curriculum occur		
Do all students on the programme take the same curriculum or are there subsets (describe)		
How is classroom contact time distributed over the calendar e.g. hours day, 1 day week		
Is PBL an addition to the to the non-PBL curriculum (Hybrid) or is PBL used in the whole curriculum		
How many and what subjects are integrated into the curriculum		
Are there subjects in the curriculum taught outside of PBL (list)		
What other teaching/learning methods are used in the curriculum (Intervention = other than PBL)		
If methods other than PBL are used how are they integrated into the PBL curriculum,		
How many students in a group/class (specifically the student teacher ratio)		
How many tutors are used with a group/class		
What are the backgrounds of the tutors (PBL – experts in substantive area of study)		
What kind of training is used to prepare the tutors/teachers		
How are the tutors/teachers evaluated and by whom		
Describe the tutorial process used by the teachers /tutors (For PBL list steps if given or shorthand e.g. Barrow's method)		
How are the problems/triggers designed: (statement of task or question given or not)		
Experiences with standardized and /or real Patients/clients incorporated into the curriculum experience		
What methods of assessment are used		
How are students graded (grades or pass/fail scores) and how are these determined.		

Reporting review results

This coding is derived from the CRD guidance, EPOC handbook and the 'Campbell Collaboration Research Design Policy Brief'. It reflects the approach to study inclusion adopted by the review group i.e. the EPOC approach. Thus it does not reflect entirely the approach that has been proposed for Campbell reviews in the above document.

Box 1: Research design reporting

i) Kind of design (free text)

ii) Randomization (RCT design only)
1. Genuinely random and concealed 2. Inadequate randomization 9.Unknown

iii) Type of comparison condition
1. No Intervention 2.Intervention as usual 3. Untargeted activity 4. Other

iv) Sample Size for Intervention group
Initial sample size for each group/sample size for this effect

v) Sample size for control group
Initial sample size for each group/sample size for this effect

vi) Similarity of control group
1. Another group from the same pool of participants 2.External 3.Archival
4. Other 9.Unknown

vii) Blinding of assessment
1. Assessor blind to status of participant or objective test 2.Assessor aware of status of participant and subjective test 9. Not known

viii) Protection from contamination
1. Intervention and control groups on the same site and/or have the same teacher
2. Intervention and control group on different sites & have different teachers
9.Unknown

ix) Baseline measures
1.Done – Groups appear evenly matched on key characteristics 2. Done - Groups do not appear evenly matched 3. Not done 9. Not known

x) Reliability
1. Done – Objective test or Kappa >0.8 2. Not Done – Not object test or Kappa <0.8
3.Not done 9. Not known

Box 2: Reporting of Description of intervention and control curricula and context of study

Country of study: (free text)

Profession or Discipline of students
1. Medicine 2. Nursing 3. Pharmacists 4. Physiotherapists

Subject of study (free text)

Academic level of course
1. Preregistraüon (not graduate level) 2. Preregistration (graduate level)
3. Masters or above 4. Certificated CPE 5. Non-certificated CPE 6 Other

Age: Mean (range)

Setting
1. University (including professional schools) 2. Workplace 3. HE College
9. Unknown

Motivation
1. No specific incentive 2. Promotion/ payrise 3. Requirement
9. Unknown

Curriculum time span (free text)

Distribution of contact time (free text)

PBL description
1. PBL sole method teaching & learning for all subjects 2. Subjects taught outside PBL
3. For intervention group PBL is an addition to curriculum 9. Unknown

Control group teaching method
1. Lecture 2. Groupwork 3. Lecture + groupwork 4. Other 9.Unknown

Tutor background
1. Subject expert 2. PBL expert 3PBL & Subject expert 9. Unknown

Student teacher ratio for intervention & control groups (free text)

Tutor Training
1. Received training and development in PBL 2. Not received training & development in PBL
9.Unknown

Tutorial process (free text)

Trigger design
1) Question/ problem given in trigger materials 2. Trigger allows free exploration

Use of patients
1. Real/ simulated patients used 2. Real simulated patients not used
9.Unknown

Method of Student assessment (free text)

Method of student grading
1. Pass/fail 2. Incremental grading system used 9.Unknown

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC Educational Resources Information Center

REPRODUCTION RELEASE

(Specific Document)

TM034895

I. DOCUMENT IDENTIFICATION:

Title: A pilot systematic review and meta-analysis on the effectiveness of Problem Based Learning	
Author(s): Mark Newman, Piet van den Bossche, Charles Engel, David Gijbels, Jean McKendree, Tony Roberts, Isobel Rolfe, John Smyth, Giovanni De Virgilio	
Corporate Source: Middlesex University (UK)	Publication Date: 2003

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1 ☑

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A ☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B ☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature:	Printed Name/Position/Title: Mr Mark Newman	
Organization/Address: Middlesex University, Furnival Building, Archway Campus, 5-10 Highgate Hill, London N19 5LW United Kingdom	Telephone: 0044 208 411 6702	FAX: 0044 208 411 6299
	E-Mail Address: m.newman@mdx.ac.uk	Date: 24/4/2003

(Over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions